

**SCHÄTZUNG VON KOVARIANZSTRUKTUREN IN
LINEAREN ZWEI- UND
MEHREBENENMODELLEN**

INAUGURAL-DISSERTATION

**zur Erlangung des Grades eines
Doktors der Wirtschafts- und Sozialwissenschaften
beim Fachbereich Philosophie und Sozialwissenschaften
der Freien Universität Berlin**

**vorgelegt von
Bernd Streitberg
aus Bad Kissingen**

1976



Digitized by the Internet Archive
in 2026

https://archive.org/details/IA41549013_0007

SCHÄTZUNG VON KOVARIANZSTRUKTUREN
IN LINEAREN ZWEI- UND MEHREBENENMODELLEN



Inaugural-Dissertation

zur Erlangung des Grades eines
Doktors der Wirtschafts- und Sozialwissenschaften
beim Fachbereich Philosophie und Sozialwissenschaften
der Freien Universität Berlin

vorgelegt von
Bernd Streitberg
aus Bad Kissingen



Tag der mündlichen Prüfung: 11.11.1976

1. Referent: Prof. Dr. Pirker

2. Referent: Prof. Dr. Kurth

VORBEMERKUNG

Die Mehrebenenanalyse ist ein relativ neues und schnell expandierendes Gebiet der empirischen Sozialforschung, das sich von der herkömmlichen Umfrageforschung durch die simultane Einbeziehung zweier oder mehrerer sozialer Aggregationsebenen unterscheidet. Die Methoden der statistischen Datenanalyse, die in mehrebenenanalytischen Studien eingesetzt werden, sind von ihren statistischen Voraussetzungen her jedoch den neuartigen Datenlagen und Fragestellungen der Mehrebenenanalyse nicht angepaßt.

Im folgenden soll für die wichtigsten Auswertungstechniken der Mehrebenenanalyse, nämlich Regressions- und Varianzanalyse (lineare Modelle), eine derartige Adaptation versucht werden. Dieser Versuch führt auf ein Modell für die Kovarianzstruktur der Beobachtungen, um die typischen Abhängigkeitsbeziehungen zwischen den Untersuchungseinheiten mehrebenenanalytischer Stichproben zu erfassen.

Legt man eine solche Kovarianzstruktur in linearen Modellen zugrunde, so hat dies Konsequenzen für die Schätzung der Modellparameter. Im Hauptteil der Arbeit geht es um die Entwicklung von Schätzverfahren für die Regressionskoeffizienten und für die Parameter der Kovarianzstruktur (durch die das Ausmaß der Abhängigkeiten zwischen den Beobachtungen beschrieben wird) sowie um die Untersuchung der finiten und asymptotischen Eigenschaften dieser Schätzfunktionen.

Das Lebenselixier wissenschaftlicher Forschung ist die kollegiale Diskussion. Für die unermüdliche Bereitschaft zu solchen Gesprächen sei stellvertretend für viele andere Gisela Arndt, Walter Bauhöfer, Wolfgang Birkenfeld, Bernd Fritzsche und Gundolf Otto gedankt.

Kenntnisreich und mit dankenswert heroischen Anstrengungen ist es Mariana Bless gelungen, die Formelwüsten des Entwurfs in eine hoffentlich nicht nur nach meiner Ansicht ästhetisch befriedigende Arbeit zu verwandeln.

INHALTSVERZEICHNIS

0. EINLEITUNG	5
0.1. Lineare Modelle in der Mehrebenenanalyse	5
0.2. Zielsetzung der Arbeit	7
0.3. Übersicht der einzelnen Kapitel	8
1. MEHREBENENSITUATIONEN UND AGGREGATIONSOPERATOREN	10
1.0. Vorbemerkung	10
1.1. Definition von Mehrebenensituationen	10
1.2. Elementare Eigenschaften von Aggregationsoperatoren	12
1.3. Simultane Diagonalisierung von Aggregationsoperatoren	16
1.4. Kovarianztheoreme	29
2. LINEARE MEHREBENENMODELLE	33
2.0. Vorbemerkung	33
2.1. Einführung des Modells	34
2.2. Interpretation der Parameter	39
2.3. OLS-Schätzer in Mehrebenenmodellen	42
2.3.1. Spezialfall: Zweiebenenmodell mit einem Regressor	42
2.3.2. Allgemeiner Fall	46
3. SCHÄTZFUNKTIONEN FÜR DIE EIGENWERTE DER KOVARIANZMATRIX	51
3.0. Vorbemerkung	51
3.1. Überblick	51
3.2. ASR-Schätzfunktionen	54

3.2.1.	Einführung	54
3.2.2.	Finite Eigenschaften	56
3.2.3.	Asymptotische Eigenschaften	65
3.3.	Generalisierte ASR-Schätzer	76
3.3.1.	Einführung	76
3.3.2.	Finite Eigenschaften	77
3.3.3.	Asymptotische Eigenschaften	78
3.4.	Lokal beste Schätzer (MINQUE)	80
3.4.1.	Einführung	80
3.4.2.	Finite Eigenschaften	84
3.4.3.	Asymptotische Eigenschaften	91
3.5.	BALESTRA-Schätzer	93
3.5.1.	Einführung	93
3.5.2.	Finite Eigenschaften	95
3.5.3.	Asymptotische Eigenschaften	102
4.	SIMULTANE SCHÄTZUNG VON EIGENWERTEN UND REGRESSIONSKOEFFIZIENTEN	103
4.1.	Feedback von ASR, MINQUE und BALESTRA-Schätzern	103
4.2.	Maximum-Likelihood-Schätzer	108
4.2.C.	Vorbemerkung	108
4.2.1.	Likelihoodfunktion und vollständig suffiziente Statistiken	108
4.2.2.	Likelihoodgleichungen und Informations- matrix	115
4.2.3.	Ein Algorithmus zur Lösung der Likeli- hoodgleichungen	117
4.2.4.	Konsistenz und asymptotische Effizienz der Maximum-Likelihood-Schätzer	122

5. VERGLEICH DER SCHÄTZFUNKTIONEN	126
5.1. Zusammenstellung der analytischen Ergebnisse	126
5.2. Skizze einiger Simulationsergebnisse	129
6. MÖGLICHE ERWEITERUNGEN	134
6.0. Vorbemerkung	134
6.1. Multivariate Regressionsmodelle	134
6.2. Testprobleme für λ und β	135
6.2.1. Test auf Unabhängigkeit der Störungen	136
6.2.2. Tests für die Regressionskoeffizienten	136
6.3. Überprüfung des Modells	139
7. LITERATURVERZEICHNIS	141

O. EINLEITUNG

O.1. LINEARE MODELLE IN DER MEHREBENENANALYSE

Die Verwendung statistischer Methoden in der Sozialforschung trägt deutlich ritualistische Züge. Statistische Auswertungstechniken werden in der Regel unverändert aus anderen Wissenschaften übernommen, ohne zu untersuchen, ob diese Verfahren den spezifischen Fragestellungen und Datenlagen der Sozialforschung gerecht werden.

Charakteristisch ist der Umgang mit linearen Modellen in empirischen Studien der Mehrebenenanalyse (MEA): Parameterschätzung und Tests vollziehen sich hier ausnahmslos im Rahmen der konventionellen Kleinstquadratetheorie (siehe BOYD [1971]), mit deren Benutzung implizit oder explizit die Annahme unabhängiger Beobachtungen verbunden ist. Ausgangspunkt der folgenden Arbeit ist die These, daß diese Unabhängigkeitsannahme in den typischen Datensituationen der Mehrebenenanalyse nicht gerechtfertigt ist.

Mehrebenenanalytische Studien unterscheiden sich von der herkömmlichen Umfrageforschung vorwiegend durch die Art des Untersuchungsdesigns.

"For the last thirty years empirical social research has been dominated by the sample survey. But as usually practiced, using random sampling of individuals, the survey is a sociological meat-grinder, tearing the individual from his social context and guar-

anteeing that nobody in the study interacts with anyone else in it. It is like a biologist putting his experimental animals through a hamburger machine and looking at every hundredth cell through a microscope: anatomy and physiology get lost, structure and function disappear, and one is left with cell biology."

(BARTON [1968])

Während in der Umfrageforschung die Unabhängigkeit der Beobachtungen artifiziell durch zufällige Auswahl aus einer großen Population hergestellt wird, wodurch mögliche Interaktionsketten zwischen den Individuen zerbrochen werden, widerspricht eine derartige Vorgehensweise der Fragestellung der Mehrebenenanalyse. Deren Ziel - nämlich aufzudecken, in welcher Weise individuelles Verhalten durch die real existierenden sozialen Kontexte, denen die Individuen angehören, beeinflusst wird - erzwingt eine möglichst vollständige Erfassung der einzelnen Mitglieder derjenigen sozialen Kontexte, die in die Untersuchung einbezogen werden. In den dichten Stichproben bzw. Vollerhebungen der MEA sind nun jedoch die Interaktionen zwischen Mitgliedern ein und desselben sozialen Kontexts nicht mehr vernachlässigbar.

Eine typische Illustration liefert ein "contextual survey" einer großen Organisation - die hierarchische formale Struktur einer derartigen Organisation induziert eine analoge Abhängigkeitsstruktur auf der Menge der einzelnen Beobachtungen in der Studie. Übersetzt man die konventionellerweise getroffene Annahme der stochastischen Unabhängigkeit dieser Beobachtungen in substanzwissenschaftliche Sprache, so entsteht die gewiß recht kühne Behauptung, daß Mitglieder derselben Arbeitsgruppe oder derselben Abteilung weder miteinander interagieren noch etwa durch gemeinsame kollektive Faktoren beein-

flußt werden.

Daß derart soziologisch unsinnige Annahmen tatsächlich in Organisationsstudien und in mehrebenenanalytischen Untersuchungen unterstellt werden, läßt sich wohl nur dadurch erklären, daß auch Standardwerke der "Statistischen Methodenlehre" für Sozialforscher sich durch erschreckende Mißverständnisse oder sogar völlige Außerachtlassung der stochastischen Voraussetzungen linearer Modelle auszeichnen (Anm. 1).

0.2. ZIELSETZUNG DER ARBEIT

In der Arbeit sollen drei Ziele erreicht werden:

- (1) Die Entwicklung eines Modells für die Kovarianzstruktur mehrebenenanalytischer Beobachtungen, das soziologisch plausibler ist als das konventionelle Unabhängigkeitsmodell. Es wird dabei von vornherein auf Situationen mit möglicherweise mehr als zwei Ebenen abgestellt.
- (2) Die Ableitung von Schätzverfahren für die beiden Parametersätze des entstehenden Modells: die Parameter des Erwartungswertteils (Regressionskoeffizienten) und die Parameter des Kovarianzteils (Eigenwerte der Kovarianzmatrix).
- (3) Die Untersuchung der finiten und asymptotischen Eigenschaften dieser Schätzfunktionen.

Anm. 1: Direkt zur Mehrebenenanalyse betrachte man etwa die Einführung linearer Modelle bei HUMMELL [1972] oder den gesamten statistischen Teil der sonst soziologisch sehr kompetenten und erfrischenden Arbeit von BOYD [1971], in der Modelle mit zufälligen Koeffizienten ohne jede Diskussion durch OLS geschätzt werden.

0.3. ÜBERSICHT DER EINZELNEN KAPITEL

Die Arbeit hat sechs Kapitel. Im ersten Kapitel werden Mehrebenensituationen mittels ihrer formalen Charakteristika definiert (Hierarchie von Beobachtungen) und eine für die Analyse derartiger Situationen wichtige Klasse linearer Abbildungen untersucht. Als "Abfallprodukt" ergibt sich ein Formalismus für die typischen Aggregations- und Disaggregationsoperationen der Mehrebenenanalyse, der das ALKER'sche Kovarianztheorem und dessen Generalisierung auf n -Ebenensituationen zu einer unmittelbaren Folgerung der elementaren Eigenschaften der betrachteten Abbildungen werden läßt.

Mit diesen Hilfsmitteln kann im zweiten Kapitel ein alternativer Ansatz für die Verteilung der Störungen in linearen Mehrebenenmodellen eingeführt und von verschiedenen Seiten her interpretiert werden. Kennzeichnend für diesen Ansatz ist die Beobachtung, daß Abhängigkeiten zwischen Mitgliedern derselben sozialen Gruppe in der Regel in kleinen Gruppen stärker sind als in großen; daß also die Stärke der Abhängigkeiten mit der Gruppengröße variiert. Für die Beurteilung bisher durchgeführter Mehrebenenstudien ist das Verhalten von OLS-Schätzern in linearen Modellen der vorgeschlagenen Struktur interessant: Abschätzungen ergeben, daß die herkömmlichen Verteilungsannahmen zu beträchtlichen Unterschätzungen der "Standardfehler" der Schätzer führen können.

Das dritte und vierte Kapitel stellt eine Reihe von Schätzfunktionen für die beiden Parametersätze des Modells vor. Im dritten Kapitel werden zunächst Schätzer für die Eigenwerte der Kovarianzmatrix entwickelt - eine Übersicht der verwendeten Strategien findet sich in Abschnitt 3.1. - und hinsichtlich ihrer finiten und asymptotischen Eigenschaften betrachtet. Eine Zusammenstellung der Annahmen, die der asymptotischen Theorie zugrunde liegen, bringt Abschnitt 3.2.3.

Im vierten Kapitel geht es um die simultane Schätzung von Eigenwerten und Regressionskoeffizienten mit zweiphasigen und Maximumlikelihood-Schätzern. Die Analyse der Likelihood-Funktion des Modells erlaubt darüber hinaus u.a. die Ableitung von notwendigen und hinreichenden Bedingungen für die vollständige Suffizienz einer bestimmten vektorwertigen Statistik. Es zeigt sich dabei, daß ein in der Praxis der MEA oft verwendetes Modell diese Bedingungen erfüllt, wodurch sich beträchtliche Vereinfachungen der Schätzverfahren ergeben.

Im fünften Kapitel werden die eingeführten Schätzfunktionen kurz verglichen und zur Illustration eine Reihe von Simulationsergebnissen dargestellt. (Anm. 1)

Das abschließende, sechste Kapitel, skizziert einige Erweiterungen der Untersuchungen des Hauptteils, die für die praktische Anwendung des Modells nötig erscheinen: die Analyse des multivariaten Regressionsmodells, die Entwicklung von Tests für Erwartungswert- und Kovarianzteil sowie mögliche Verfahren zur Überprüfung des Modells selbst.

1

Anm.1: Auf die Anwendung des Modells auf reale Daten wurde zunächst verzichtet, da am Institut für Soziologie zur Zeit eine Diplomarbeit von Walter Bauhöfer entsteht, die unter anderem derartige Anwendungen zum Inhalt hat.

1. MEHREBENENSITUATIONEN UND AGGREGATIONSOPERATOREN

1.0. VORBEMERKUNG

Im ersten Kapitel geht es vor allem darum, eine Klasse linearer Abbildungen einzuführen, mit deren Hilfe die Aggregations- und Disaggregationsoperationen der Mehrebenenanalyse bequem darzustellen sind. Als Voraussetzung werden dazu zunächst die Eigenschaften typischer Datensituationen der Mehrebenenanalyse formal (und ziemlich trocken) beschrieben.

1.1. DEFINITION VON MEHREBENENSITUATIONEN

Mehrebenensituationen sollen hier über Zerlegungen einer gegebenen Grundmenge von Objekten definiert werden. In den Anwendungen der empirischen Sozialforschung handelt es sich bei den Elementen der Grundmenge um beobachtete menschliche Individuen - etwa um die Mitglieder einer Organisation; den Mengen einer Zerlegung entsprechen soziale Aggregate wie Abteilungen oder Bereiche.

$\mathcal{M}$ sei eine nichtleere Menge. Eine Zerlegung $\mathcal{Z}$ von $\mathcal{M}$ in m Teilmengen ist eine Klasse $\{\mathcal{M}_1, \mathcal{M}_2, \dots, \mathcal{M}_m\}$, wobei

$$(1) \quad \mathcal{M}_j \subseteq \mathcal{M}, \quad \mathcal{M}_j \neq \emptyset; \quad j = 1..m$$

$$(2) \quad \mathcal{M}_i \cap \mathcal{M}_j = \emptyset \quad i, j = 1..m; i \neq j$$

$$(3) \quad \bigcup_{j=1}^m \mathcal{M}_j = \mathcal{M}.$$

Eine Zweiebenensituation ist ein Paar $(\mathcal{U}, \mathcal{Z})$, wobei

- (1) $\mathcal{U}$ eine nichtleere, endliche Menge
- (2) $\mathcal{Z}$ eine Zerlegung von $\mathcal{U}$ in m Teilmengen
 $\{\mathcal{U}_1, \mathcal{U}_2 \dots \mathcal{U}_m\}$

ist.

Die Teilmengen $\mathcal{U}_j$ ($j=1..m$) einer Zerlegung bezeichnen wir im Zusammenhang der Mehrebenenanalyse als Aggregate, Gruppen oder Kollektive.

Der Begriff "Zweiebenensituation" rührt daher, daß wir $\mathcal{U}$ und $\mathcal{Z}$ als Mengen auffassen können, deren Elemente zwei verschiedenen Aggregationsebenen angehören: der Mikroebene (die aus den Elementen von $\mathcal{U}$ besteht) und der Makroebene (den Aggregaten $\mathcal{U}_1 \dots \mathcal{U}_m$ in $\mathcal{Z}$).

Es wäre möglich, eine n -Ebenensituation analog als n -tupel $(\mathcal{U}, \mathcal{Z}_1, \mathcal{Z}_2 \dots \mathcal{Z}_{n-1})$ mit $\mathcal{Z}_j$ Zerlegung von $\mathcal{U}$ zu definieren. Stellt $\mathcal{U}$ etwa die Menge aller Befragten einer Stadt dar, so könnte $\mathcal{Z}_1$ die Zerlegung dieser Menge in Familien, $\mathcal{Z}_2$ die Zerlegung in Arbeitsstätten, $\mathcal{Z}_3$ die Zerlegung in regionale Kontexte bilden.

Eine derartige Definition ist für unsere Zwecke jedoch zu allgemein - man untersucht in der herkömmlichen Mehrebenenanalyse ausschließlich Systeme, die sich als Hierarchie sozialer Aggregate beschreiben lassen: Arbeitsgruppen, Abteilungen, Bereiche, Betriebe in Organisationsstudien; Nachbarschaften, Regionen, Bezirke, Städte in Untersuchungen der Stadtforschung. Zwei Zerlegungen der Objektmenge stehen hier nicht beliebig nebeneinander, sondern die eine Zerlegung bildet stets eine Verfeinerung oder Vergröberung der anderen.

Die Zerlegung $\mathcal{Z} = \{\mathcal{U}_1, \mathcal{U}_2 \dots \mathcal{U}_m\}$ von $\mathcal{U}$ heißt dabei Verfeinerung einer Zerlegung $\mathcal{Z}' = \{\mathcal{U}'_1, \mathcal{U}'_2 \dots \mathcal{U}'_m\}$ von $\mathcal{U}$, wenn für alle $\mathcal{U}_i$ in $\mathcal{Z}$ ein $\mathcal{U}'_j$ in $\mathcal{Z}'$ existiert, so daß

$\mathcal{U}_i \subseteq \mathcal{U}_j$. Wir nennen $\mathcal{Z}'$ dann auch Vergrößerung von $\mathcal{Z}$ und schreiben $\mathcal{Z} < \mathcal{Z}'$, sofern $\mathcal{Z}$ Verfeinerung von $\mathcal{Z}'$ und $\mathcal{Z} \neq \mathcal{Z}'$.

Eine n-Ebenensituation ist ein n-tupel ($n \geq 2$)

$(\mathcal{U}, \mathcal{Z}_1, \mathcal{Z}_2 \dots \mathcal{Z}_{n-1})$, wobei

- (1) $\mathcal{U}$ eine nichtleere, endliche Menge
- (2) $\mathcal{Z}_i$ ($i=1..n-1$) Zerlegung von $\mathcal{U}$ ist;
- (3) für $i < j$ gilt, daß $\mathcal{Z}_i < \mathcal{Z}_j$ ($i, j=1..n-1$).

Die Zerlegungen, durch die eine n-Ebenensituation definiert wird, stehen also in einer hierarchischen Ordnung:

$$\mathcal{Z}_1 < \mathcal{Z}_2 < \dots < \mathcal{Z}_{n-2} < \mathcal{Z}_{n-1}.$$

$\mathcal{Z}_1$ ist die feinste, $\mathcal{Z}_{n-1}$ die grösste Zerlegung in $(\mathcal{Z}_1, \mathcal{Z}_2 \dots \mathcal{Z}_{n-1})$.

1.2. ELEMENTARE EIGENSCHAFTEN VON AGGREGATIONSOPERATOREN

Wir wählen eine Anordnung $(a_1, a_2 \dots a_N)$ der Elemente in $\mathcal{U}$ sowie für die Zerlegung $\mathcal{Z}_i$ der Mehrebenensituation eine Anordnung $(\mathcal{Z}_{i1}, \mathcal{Z}_{i2} \dots \mathcal{Z}_{im_i})$ der m_i Aggregate dieser Zerlegung.

Es können dann für jede Zerlegung $\mathcal{Z}_i$ zwei Matrizen G_i bzw. Δ_i definiert werden, die wir als Inzidenzmatrix bzw. als Aggregationsmatrix der Zerlegung bezeichnen.

Die Inzidenzmatrix G_i der Zerlegung $\mathcal{Z}_i = (\mathcal{Z}_{i1} \dots \mathcal{Z}_{im_i})$ ist eine (N, m_i) -Matrix $G_i = (g_{jk}^{(i)})$ mit

$$g_{jk}^{(i)} := \begin{cases} 1 & \text{für } a_j \in \mathcal{Z}_{ik} \\ 0 & \text{für } a_j \notin \mathcal{Z}_{ik} \end{cases}.$$

Bezeichnet n_{ik} die Anzahl der Elemente in $\mathcal{Z}_{ik}$, so gilt

$$G_i^t G_i = \text{diag}[n_{i1}, n_{i2}, \dots, n_{im_i}] = \begin{bmatrix} n_{i1} & & & & \\ & n_{i2} & & & \\ & & \ddots & & \\ & & & \ddots & \\ & & & & n_{im_i} \end{bmatrix}$$

Die Aggregationsmatrix Δ_i der Zerlegung $\mathcal{Z}_i$ ist eine (N, N) -Matrix

$$\Delta_i := G_i (G_i^t G_i)^{-1} G_i^t .$$

Ist $d_{st}^{(i)}$ das Element in der Zeile s und Spalte t von Δ_i , so gilt

$$d_{st}^{(i)} = \begin{cases} \frac{1}{n_{ik}} & \text{falls } a_s \in \mathcal{Z}_{ik} \text{ und } a_t \in \mathcal{Z}_{ik} \\ 0 & \text{sonst} \end{cases}$$

BEISPIEL: Sei $\mathcal{U} = (1, 2, 3, 4, 5)$, $\mathcal{Z}_1 = (\{1, 2, 5\}, \{3, 4\})$, so ist

$$G_1 = \begin{bmatrix} 1 & 0 \\ 1 & 0 \\ 0 & 1 \\ 0 & 1 \\ 1 & 0 \end{bmatrix} \quad \Delta_1 = \begin{bmatrix} 1/3 & 1/3 & 0 & 0 & 1/3 \\ 1/3 & 1/3 & 0 & 0 & 1/3 \\ 0 & 0 & 1/2 & 1/2 & 0 \\ 0 & 0 & 1/2 & 1/2 & 0 \\ 1/3 & 1/3 & 0 & 0 & 1/3 \end{bmatrix}$$

Die Interpretation dieser Matrizen gewinnt an Farbe, wenn wir den $\mathbb{R}^N$ bzw. den $\mathbb{R}^m$ mit Basen (etwa den Standardbasen) ausstatten und die linearen Abbildungen untersuchen, die von Inzidenz- bzw. Aggregationsmatrizen dargestellt werden. Die Vektoren des $\mathbb{R}^N$ sollen dabei Meßwerte für die N betrachteten Individuen enthalten, die Vektoren des $\mathbb{R}^m$ Meßwerte für die

m_i Aggregate der i 'ten Zerlegung enthalten.

Sei $x \in \mathbb{R}^{m_i}$, dann ordnet $G_i x$ jedem Individuum in $\mathcal{M}$ den Wert des Kollektivs zu, dem dieses Individuum zugehört (Disaggregation); $(G_i' G_i)^{-1} G_i' y$ liefert für jedes Aggregat in $\mathcal{Z}_i$ das arithmetische Mittel seiner Mitgliedsobjekte in $y \in \mathbb{R}^N$ (Bildung von Gruppendurchschnitten) und $\Delta_i y = G_i (G_i' G_i)^{-1} G_i' y$ weist diese Gruppendurchschnitte wieder den Individuen in $\mathcal{M}$ zu (Aggregation).

Die Entscheidung, Aggregationen nicht als Abbildungen des $\mathbb{R}^N$ in den $\mathbb{R}^{m_i}$, sondern als Disaggregation von Gruppendurchschnitten aufzufassen, ermöglicht es, die Wirkung des Hintereinanderschaltens von Aggregationen nach möglicherweise verschiedenen Zerlegungen stets im gleichen $\mathbb{R}^N$ zu studieren.

Die folgenden Sätze sind einfach, aber grundlegend. Sie besagen, anschaulich gesprochen, daß ein Vektor (des $\mathbb{R}^N$), der bereits nach $\mathcal{Z}_i$ aggregiert wurde, durch weitere Aggregation nach $\mathcal{Z}_i$ oder einer noch feineren Zerlegung nicht verändert wird.

SATZ 1.2.1. : Sei Δ_i Aggregationsmatrix einer Zerlegung $\mathcal{Z}_i$ von $\mathcal{M}$ in m_i Aggregate. Dann gilt

- (1) Δ_i ist Projektor, also eine idempotente ($\Delta_i^2 = \Delta_i$) und symmetrische Matrix.
- (2) $\text{tr}(\Delta_i) = \text{rg}(\Delta_i) = m_i$ (Anm. 1)

BEWEIS:

- (1) ergibt sich direkt aus der Definition:

$$\begin{aligned} \Delta_i^2 &= G_i (G_i' G_i)^{-1} G_i' G_i (G_i' G_i)^{-1} G_i' \\ &= G_i (G_i' G_i)^{-1} G_i' \\ &= \Delta_i \end{aligned}$$

Anm. 1: Mit $\text{rg} A$ sei der Rang, mit $\text{tr} A$ die Spur einer Matrix A bezeichnet.

- (2) Da Δ_i symmetrisch ist, existiert eine orthogonale Matrix H , die Δ_i auf Diagonalform transformiert:

$$H' \Delta_i H = \Lambda_i$$

Wegen Idempotenz von Δ_i enthält Λ_i auf der Hauptdiagonale nur die Werte 0 und 1. Man findet also

$$\text{rg } \Delta_i = \text{rg } \Lambda_i = \text{tr } \Lambda_i = \text{tr } H' \Delta_i H = \text{tr } \Delta_i \quad \text{und}$$

$$\text{tr } \Delta_i = \text{tr } (G_i' G_i)^{-1} G_i' G_i = m_i.$$

Der Eigenraum von Δ_i (also derjenige Unterraum des $\mathbb{R}^N$, auf den Δ_i projiziert) hat also die Dimension m_i ; er besteht aus allen Vektoren des $\mathbb{R}^N$, die "gruppenweise konstant" sind (Anm. 1).

Der Kern von Δ_i (also derjenige Unterraum des $\mathbb{R}^N$, den Δ_i auf den Nullvektor abbildet) hat folglich die Dimension $N - m_i$; er besteht aus allen Vektoren des $\mathbb{R}^N$ mit "verschwindenden Gruppenmitteln".

SATZ 1.2.2. : Seien Δ_i, Δ_j Aggregationsmatrizen der Zerlegung $\mathcal{Z}_i$ und $\mathcal{Z}_j$ von $\mathcal{U}$, wobei $\mathcal{Z}_i < \mathcal{Z}_j$. Dann gilt

$$\Delta_i \Delta_j = \Delta_j \Delta_i = \Delta_j.$$

BEWEIS : Zunächst gilt $\mathcal{R}(\Delta_j) \subseteq \mathcal{R}(\Delta_i)$ (Anm. 2):
Denn sei $x \in \mathcal{R}(\Delta_j)$, dann ist x $\mathcal{Z}_{jS}$ -konstant für alle

Anm. 1: Sei $\mathcal{P} \subseteq \mathcal{U}$, $\mathcal{P} = \{a_{p_1}, a_{p_2}, \dots, a_{p_r}\}$. Dann heiße $x \in \mathbb{R}^N$ $\mathcal{P}$ -konstant, wenn $x_{p_1} = x_{p_2} = \dots = x_{p_r}$. Der Eigenraum von Δ_i besteht also aus allen Vektoren x , die $\mathcal{Z}_{ik}$ -konstant sind für alle $\mathcal{Z}_{ik} \in \mathcal{Z}_i$.

Anm. 2: $\mathcal{R}(A)$ bezeichne den Spaltenraum der Matrix A , d.h. wenn A vom Typ (a, b) ist, die Menge aller Vektoren $z \in \mathbb{R}^a$ mit $z = Ax$ für $x \in \mathbb{R}^b$. $\mathcal{N}(A)$ bezeichne den Kern der Matrix A .

$\beta_{js} \in \beta_j$ und damit auch β -konstant für alle $\beta \subseteq \beta_{js}$.
 Wegen $\beta_i < \beta_j$ ist x also speziell für alle $\beta_{ik} \in \beta_i$
 konstant und liegt damit im Eigenraum von Δ_i .
 Wir zeigen $\Delta_i \Delta_j = \Delta_j$ (die Behauptung $\Delta_i \Delta_j = \Delta_j$ folgt daraus
 durch Transposition):
 Jeder Vektor $z \in \mathbb{R}^N$ besitzt die eindeutige Zerlegung
 $z = z_1 + z_0$ mit $z_1 \in \mathcal{R}(\Delta_j)$ und $z_0 \in \mathcal{N}(\Delta_j)$. Es ist
 $\Delta_j z = z_1$. Wegen $z_1 \in \mathcal{R}(\Delta_j) \subseteq \mathcal{R}(\Delta_i)$ gilt $\Delta_i z_1 = z_1$,
 also $\Delta_i \Delta_j z = \Delta_j z$ für alle $z \in \mathbb{R}^N$.

1.3. SIMULTANE DIAGONALISIERUNG VON AGGREGATIONSOPERATOREN

Alle Aggregationsoperatoren einer Mehrebenensituation kommutieren -
 es muß folglich möglich sein, eine orthogonale Matrix T so zu
 finden, daß T gleichzeitig alle Aggregationsoperatoren Δ_i
 diagonalisiert, d.h. daß $T' \Delta_i T$ für alle $i=1 \dots n-1$ eine Dia-
 gonalmatrix darstellt. Im folgenden geht es darum, eine der-
 artige Matrix T explizit zu konstruieren.

Um die Eigenschaften von Aggregationsoperatoren im Zusammenhang
 darstellen zu können, studieren wir dieses Problem bereits hier -
 benötigt werden die Ergebnisse erst später.

Die Strategie, der wir folgen wollen, sei kurz erläutert - sie
 beruht auf drei Ideen:

- (1) Jede Matrix T mit der gewünschten Eigenschaft diagonalisiert mit $\Delta_1, \Delta_2, \dots, \Delta_{n-1}$ auch alle Linearkombinationen

$$\Sigma = c_0 I + c_1 \Delta_1 + \dots + c_{n-1} \Delta_{n-1}$$

der Matrizen $I, \Delta_1, \dots, \Delta_{n-1}$. Andererseits leistet jede
 Matrix, die Σ unabhängig von der speziellen Wahl der
 c_i in Diagonalform überführt, das Gewünschte.

(2) Die Spektraldarstellung von Σ :

$$\Sigma = \lambda_0 K_0 + \lambda_1 K_1 + \dots \lambda_{n-1} K_{n-1}$$

kann auf einfache Weise explizit angegeben werden. Dabei sind die K_i idempotente Matrizen mit $K_i K_j = 0$ für $i \neq j$, die nicht von den c_i abhängen; λ_i sind die Eigenwerte von Σ .

(3) Das Problem zerfällt damit in n Teilprobleme, nämlich die Suche nach Orthogonalbasen $T_0, T_1, \dots, T_{n-1}$ für die Spaltenräume von $K_0, K_1, \dots, K_{n-1}$. Diese Teilprobleme sind unabhängig voneinander zu lösen, da mit den Spaltenräumen der K_i auch Basen für diese Spaltenräume wechselseitig orthogonal sind. Die gesuchte Matrix T ist daher einfach $[T_0 | T_1 | \dots | T_{n-1}]$.

Der Übergang zu Linearkombinationen benötigt keinen Beweis.

Der Term $c_0 I$ wird dabei hinzugenommen, da T auch die Eigenvektoren zum Eigenwert 0 der Δ_i enthalten soll (Σ spannt im allgemeinen den ganzen $\mathbb{R}^N$ auf).

SATZ 1.3.1. : Sei $\Sigma = c_0 I + c_1 \Delta_1 + \dots c_{n-1} \Delta_{n-1}$ mit $c_i \in \mathbb{R}$ ($i = 0 \dots n-1$).

(1) Σ besitzt die Darstellung

$$\Sigma = \lambda_0 K_0 + \lambda_1 K_1 + \dots \lambda_{n-1} K_{n-1}$$

mit

$$\lambda_i = c_0 + c_1 + \dots c_i \quad (i = 0 \dots n-1)$$

und

$$K_i = \begin{cases} (I - \Delta_1) & \text{für } i = 0 \\ (\Delta_i - \Delta_{i+1}) & \text{für } i = 1 \dots n-2 \\ \Delta_{n-1} & \text{für } i = n-1 \end{cases}$$

- (2) Die Eigenwerte von Σ sind λ_i ($i = 0 \dots n-1$), die Hauptidempotenten von Σ sind die Matrizen K_i ($i = 0 \dots n-1$) - es gilt also

$$K_i K_j = \begin{cases} K_i & \text{für } i = j \\ 0 & \text{für } i \neq j \end{cases} \quad i, j = 0 \dots n-1$$

$$I = K_0 + K_1 + \dots K_{n-1}$$

$$\Sigma K_i = \lambda_i K_i$$

BEWEIS

- (1) Die Menge aller Σ mit $\Sigma = c_0 I + c_1 \Delta_1 + \dots c_{n-1} \Delta_{n-1}$ stellt einen Unterraum des Vektorraums $\mathbb{R}^{N,N}$ aller N -reihigen Matrizen dar.

Der Vektor $c = (c_0, c_1, \dots, c_{n-1})$ ist die Koordinatendarstellung von Σ bezüglich der Basis $(I, \Delta_1, \dots, \Delta_{n-1})$. Wir betrachten die Basistransformation

$$B : (I, \Delta_1, \dots, \Delta_{n-1}) \mapsto (K_0, K_1, \dots, K_{n-1}) ,$$

die Zeilen der Matrix dieser Abbildung erhalten wir durch Darstellung der Elemente der zweiten Basis als Linearkombinationen der ersten, es gilt

$$B = \begin{bmatrix} 1 & -1 & 0 & \dots & 0 \\ 0 & 1 & -1 & \dots & 0 \\ . & & & & \\ . & & & & \\ 0 & 0 & 0 & \dots & 1 \end{bmatrix}$$

Der Vektor $\lambda = (\lambda_0, \dots, \lambda_{n-1})$, der die Koordinaten von Σ bezüglich der Basis $(K_0, \dots, K_{n-1})$ enthält, ist mit c über

$$c' = \lambda' B$$

oder

$$\lambda = (B')^{-1} c$$

verknüpft. Man findet rekursiv

$$(B')^{-1} = \begin{bmatrix} 1 & & & & \\ & 1 & & & \\ & & 1 & & \\ & & & 1 & \\ & & & & \ddots \\ & & & & & 1 \\ 1 & 1 & 1 & \dots & 1 \end{bmatrix}$$

- (2) Mit Δ_1 ist auch $K_0 = (I - \Delta_1)$ idempotent. Es gilt für $i = 1, 2, \dots, n-2$

$$K_i^2 = (\Delta_i - \Delta_{i+1}) (\Delta_i - \Delta_{i+1}) = \Delta_i - \Delta_{i+1} - \Delta_{i+1} + \Delta_{i+1} = K_i.$$

Analog prüft man $K_i K_j = 0$ und $K_0 + K_1 + \dots + K_{n-1} = I$ durch Rechnung nach.

Daß λ_i Eigenwert von Σ ist, ergibt sich dann aus

$$\Sigma K_i = (\lambda_0 K_0 + \dots + \lambda_1 K_1 + \dots + \lambda_{n-1} K_{n-1}) K_i = \lambda_i K_i^2 = \lambda_i K_i.$$

Orthogonalbasen für die Spaltenräume von $K_0 = (I - \Delta_1)$ und $K_{n-1} = \Delta_{n-1}$ findet man, wenn man die Eigenzerlegung einer allgemeinen Aggregationsmatrix Δ angibt. Die Eigenvektoren zum Eigenwert 0 von Δ_1 spannen dann den Spaltenraum $\mathcal{R}(K_0)$ von K_0 auf, die Eigenvektoren zum Eigenwert 1 von Δ_{n-1} stellen eine Orthogonalbasis für K_{n-1} dar. Die Eigenvektoren sind nicht eindeutig bestimmt, eine gut zu interpretierende (und leicht zu berechnende) Basis der m - bzw. $(N-m)$ -dimensionalen Eigenräume von Δ zum Eigenwert 1 bzw. 0 lässt sich mit Hilfe von HELMERT-Matrizen gewinnen (siehe zu HELMERT-Matrizen etwa LANCASTER [1965]).

Wir beginnen mit einer Definition:

Für jede natürliche Zahl $k > 1$ sei als HELMERT-Matrix H_k

der Ordnung k eine (k,k) -Matrix $H_k = (h_{ij}^{(k)})$ bezeichnet, für

die

$$h_{ij}^{(k)} = \begin{cases} 0 & \text{für } i > j + 1 \\ -j & \text{für } i = j + 1 \\ 1 & \text{für } i < j + 1 \end{cases}$$

BEISPIELE (für $k=1,2,3,4,5$) :

$$H_1 = [1]$$

$$H_2 = \begin{bmatrix} 1 & 1 \\ -1 & 1 \end{bmatrix}$$

$$H_3 = \begin{bmatrix} 1 & 1 & 1 \\ -1 & 1 & 1 \\ 0 & -2 & 1 \end{bmatrix}$$

$$H_4 = \begin{bmatrix} 1 & 1 & 1 & 1 \\ -1 & 1 & 1 & 1 \\ 0 & -2 & 1 & 1 \\ 0 & 0 & -3 & 1 \end{bmatrix}$$

$$H_5 = \begin{bmatrix} 1 & 1 & 1 & 1 & 1 \\ -1 & 1 & 1 & 1 & 1 \\ 0 & -2 & 1 & 1 & 1 \\ 0 & 0 & -3 & 1 & 1 \\ 0 & 0 & 0 & -4 & 1 \end{bmatrix}$$

Man erkennt, daß je zwei verschiedene Spalten von H_k orthogonal sind und daß $H_k' H_k$ eine (k,k) -Diagonalmatrix

$$H_k' H_k = \text{diag}[1.2, 2.3, 3.4, \dots, k(k+1), k]$$

darstellt.

Wir wollen uns im folgenden die Elemente von $\mathcal{M}$ so angeordnet denken, daß jede Aggregationsmatrix Δ_i der Mehrebenensituation blockdiagonale Gestalt hat. Ausgehend von einer beliebigen Anordnung der Aggregate in $\mathcal{Z}_1, \dots, \mathcal{Z}_{n-1}$ ist es stets möglich, alle Matrizen Δ_i auf blockdiagonale Form zu bringen: man "sortiert" die Elemente von $\mathcal{M}$ "stabil" fortlaufend nach $\mathcal{Z}_1, \mathcal{Z}_2 \dots \mathcal{Z}_{n-1}$. Stabiles Sortieren nach $\mathcal{Z}_i$ bedeutet dabei, daß die jeweils vorgefundene Anordnung aller Elemente eines Aggregats in $\mathcal{Z}_i$ untereinander durch das Sortieren nicht zerstört wird.

Sei nun $\mathcal{Z}_i$ eine Zerlegung von $\mathcal{M}$ in Aggregate $(\mathcal{Z}_{i1}, \mathcal{Z}_{i2}, \dots, \mathcal{Z}_{im})$ mit $n_1, n_2, \dots, n_m$ Elementen und Δ_i die entsprechende blockdiagonale Aggregationsmatrix. Δ_i ist also die direkte Summe von Matrizen der Form $\frac{1}{n_j} \Pi_{n_j} \Pi'_{n_j}$ (Anm. 1):

$$\Delta_i = \bigoplus_{j=1}^m \left(\frac{1}{n_j} \Pi_{n_j} \Pi'_{n_j} \right)$$

Wir definieren eine Matrix H vom Typ (N, N) durch

$$H = \bigoplus_{j=1}^m H_{n_j} = \begin{bmatrix} H_{n_1} & & & \\ & H_{n_2} & & \\ & & \ddots & \\ \emptyset & & & H_{n_m} \end{bmatrix}$$

Anm. 1: Mit Π_k bezeichnen wir einen Vektor $(1 \ 1 \ \dots \ 1)'$, der k Einsen enthält, mit $\emptyset_k$ einen Vektor mit k Nullen - sofern die Dimension der Vektoren aus dem Zusammenhang hervorgeht, schreiben wir kurz Π bzw. $\emptyset$.

Die direkte Summe zweier Matrizen A und B ist die Matrix

$$A \oplus B := \begin{bmatrix} A & \emptyset \\ \emptyset & B \end{bmatrix}$$

Sofern die auftretenden Operationen definiert sind, gelten folgende Eigenschaften (MACDUFFEE[1946]):

- (1) $(A \oplus B) \oplus C = A \oplus (B \oplus C)$
- (2) $(A \oplus B) + (C \oplus D) = (A+B) \oplus (C+D)$
- (3) $(A \oplus B)(C \oplus D) = AC \oplus BD$
- (4) $(A \oplus B)' = A' \oplus B'$
- (5) $(A \oplus B)^{-1} = A^{-1} \oplus B^{-1}$
- (6) $\det(A \oplus B) = \det(A) \det(B)$

Für $A_1 \oplus A_2 \oplus \dots \oplus A_n$ schreiben wir auch $\bigoplus_{i=1}^n A_i$.

und zeigen, daß die Spalten dieser direkten Summe von "HELMERT-Blöcken" ein vollständiges System von zueinander orthogonalen Eigenvektoren von Δ_i liefern.

SATZ 1.3.2. : Die Matrix $H(H'H)^{-1/2}$ ist orthogonal und es gilt $(H'H)^{-1/2} H' \Delta_i H(H'H)^{-1/2} = \Delta_i$ mit

$$\Delta_i = \bigoplus_{j=1}^m \Delta_{n_j}^{(i)} \quad \text{wobei} \quad \Delta_{n_j}^{(i)} = \text{diag}[0, 0, \dots, 0, 1] \quad \text{vom Typ} \\ (n_j, n_j) .$$

BEWEIS : Wegen der blockdiagonalen Struktur der Matrizen H und Δ_i ist es nur nötig, den Beweis für einen Diagonalblock von Δ_i : $(n_j)^{-1} \Pi \Pi'$ und den entsprechenden Diagonalblock H_{n_j} von H zu führen.

Offenbar ist Π orthogonal zu den ersten (n_{j-1}) Spalten von H_{n_j} und das Skalarprodukt von Π mit der letzten Spalte von H_{n_j} ist gleich n_j , so daß

$$\Pi' H_{n_j} = (0 \ 0 \ \dots \ 0 \ n_j)$$

und

$$(n_j)^{-1} H_{n_j}' \Pi \Pi' H_{n_j} = \begin{bmatrix} 0 & & & & \\ & 0 & & & \\ & & 0 & & \\ & & & \ddots & \\ & & & & 0 \\ \emptyset & & & & & n_j \end{bmatrix} ,$$

woraus die Behauptung folgt.

Satz 1.3.2. löst zwei der Teilprobleme, auf die sich unsere Suche nach einer orthogonalen Matrix, die alle Δ_i diagonalisiert, reduziert hatte: die Konstruktion von Orthogonalbasen

für K_0 und K_{n-1} .

Wir finden T_{n-1} , wenn wir in allen Diagonalblöcken der Matrix H mit den Eigenvektoren von Δ_{n-1} die letzte Spalte herausziehen - dies gibt G_{n-1} - und normieren.

Wir finden T_0 , wenn wir in allen Diagonalblöcken der Matrix H mit den Eigenvektoren von Δ_1 die letzte Spalte streichen und die verbleibenden $(N-m_1)$ orthogonalen Spalten normieren.

BEISPIEL: Sei $\mathcal{M} = (1,2,3,4,5,6,7,8,9)$;

$\mathcal{Z}_1 = (\{1,2\}, \{3,4,5\}, \{6,7,8\}, \{9\})$; $\mathcal{Z}_{n-1} := (\{1,2,3,4,5\}, \{6,7,8,9\})$

Dann ergibt sich mit Satz 1.3.2.:

$$T_0 = \begin{bmatrix} 1 & 0 & 0 & 0 & 0 \\ -1 & 0 & 0 & 0 & 0 \\ 0 & 1 & 1 & 0 & 0 \\ 0 & -1 & 1 & 0 & 0 \\ 0 & 0 & -2 & 0 & 0 \\ 0 & 0 & 0 & 1 & 1 \\ 0 & 0 & 0 & -1 & 1 \\ 0 & 0 & 0 & 0 & -2 \\ 0 & 0 & 0 & 0 & 0 \end{bmatrix} \cdot \begin{bmatrix} 1/\sqrt{2} & & & & \\ & 1/\sqrt{2} & & & \\ & & \emptyset & & \\ & & & 1/\sqrt{6} & \\ & & & & 1/\sqrt{2} \\ \emptyset & & & & & 1/\sqrt{6} \end{bmatrix}$$

$$T_{n-1} = \begin{bmatrix} 1 & 0 \\ 1 & 0 \\ 1 & 0 \\ 1 & 0 \\ 1 & 0 \\ 0 & 1 \\ 0 & 1 \\ 0 & 1 \\ 0 & 1 \end{bmatrix} \cdot \begin{bmatrix} 1/\sqrt{5} & \\ & 1/\sqrt{4} \end{bmatrix}$$

Die Spalten von T_0 definieren orthogonale Kontraste zwischen den Beobachtungen innerhalb eines Aggregats von $\mathcal{Z}_1$, und zwar werden in einem Aggregat mit k Elementen jeweils die ersten

j ($j=1,2,\dots,k-1$) Beobachtungen des Aggregats mit der $(j+1)$ 'ten Beobachtung verglichen.

Es bleibt das Problem, Eigensysteme zum Eigenwert 1 für $\Delta_i - \Delta_{i+1}$ ($i=1,2,\dots,n-2$) zu finden. Die Rolle der Transformationen H_k wird dabei von Matrizen übernommen, die analog zu HELMERT-Matrizen aufgebaut sind und als "HELMERT-Matrizen zweiter Ordnung" bezeichnet werden können: vermittelte H_k Kontraste zwischen den Elementen eines Aggregats, so definieren diese Transformationen Kontraste zwischen den Unteraggregaten eines Aggregats.

Seien Δ_i und Δ_{i+1} die blockdiagonalen Aggregationsmatrizen von $\mathcal{Z}_i$ und $\mathcal{Z}_{i+1}$, wobei $\mathcal{Z}_i$ eine Zerlegung von $\mathcal{M}$ in m_i Aggregate und $\mathcal{Z}_{i+1}$ eine Zerlegung von $\mathcal{M}$ in m_{i+1} Aggregate ist. Betrachten wir den Diagonalblock von Δ_{i+1} , der zum Aggregat $\mathcal{Z}_{(i+1),r}$ ($r=1,2,\dots,m_{i+1}$) von $\mathcal{Z}_{i+1}$ gehört. Die entsprechende Untermatrix von Δ_i besteht aus p Diagonalblöcken $\frac{1}{k_j} \mathbb{1} \mathbb{1}'$, deren jeder einem Unteraggregat von $\mathcal{Z}_{(i+1),r}$ mit k_j ($j=1,2,\dots,p$) Elementen entspricht. Es gilt $k_1 + k_2 + \dots + k_p = n_r$, wenn $\mathcal{Z}_{(i+1),r}$ n_r Elemente umfaßt.

Wir definieren als HELMERT-Matrix zweiter Ordnung eine (n_r, p) -Matrix

$$P_r := \begin{bmatrix} \mathbb{1}_{k_1} & \mathbb{1}_{k_1} & \cdot & \cdot & \cdot & \mathbb{1}_{k_1} & \mathbb{1}_{k_1} \\ -\varepsilon_1 \mathbb{1}_{k_2} & \mathbb{1}_{k_2} & \cdot & \cdot & \cdot & \mathbb{1}_{k_2} & \mathbb{1}_{k_2} \\ 0_{k_3} & -\varepsilon_2 \mathbb{1}_{k_3} & \cdot & \cdot & \cdot & \mathbb{1}_{k_3} & \mathbb{1}_{k_3} \\ \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot \\ 0_{k_p} & 0_{k_p} & \cdot & \cdot & \cdot & -\varepsilon_{p-1} \mathbb{1}_{k_p} & \mathbb{1}_{k_p} \end{bmatrix}$$

wobei

$$g_j := (k_{j+1})^{-1} \sum_{t=1}^j k_t \quad (j=1, 2 \dots p-1)$$

$\bar{P}$ sei die direkte Summe der Matrizen $P_1, P_2 \dots P_r \dots P_{m_{i+1}}$:

$$\bar{P} = \bigoplus_{r=1}^{m_{i+1}} P_r = \begin{bmatrix} P_1 & & & & \\ & P_2 & & & \\ & & \ddots & & \\ & & & P_r & \\ & & & & \ddots \\ & \emptyset & & & & P_{m_{i+1}} \end{bmatrix}$$

und P sei die $(N, m_i - m_{i+1})$ -Matrix, die aus $\bar{P}$ durch Streichen der jeweils letzten Spalte in allen Blöcken P_r entsteht.

Der nächste Satz zeigt, daß $T_i := P(P'P)^{-1/2}$ die gesuchte Matrix der Eigenvektoren zum Eigenwert 1 von $(\Delta_i - \Delta_{i+1})$ darstellt.

SATZ 1.3.3. :

(1) Die Spalten der Matrix $\bar{P}$ sind orthogonal.

(2) $\text{rg}(T_i) = \text{rg}(P) = \text{rg}(\Delta_i - \Delta_{i+1}) = m_i - m_{i+1}$

(3) $(\Delta_i - \Delta_{i+1}) T_i = T_i$

BEWEIS :

(1) Bezeichnen wir die Spalten von P_r mit $h_1, h_2 \dots h_p$, so ist

$$\begin{aligned} h'_p h_s &= \eta'_s h_s = k_1 + k_2 + \dots + k_s - g_s k_{s+1} \\ &= g_s k_{s+1} - g_s k_{s+1} = 0 \quad (s = 1, 2 \dots p-1) \end{aligned}$$

und für $s < t$ ($s, t = 1, 2 \dots p-1$)

$$h'_s h_t = h'_s \eta = 0.$$

Für $s=t$ gilt

$$h_s' h_s = k_1 + k_2 + \dots + k_s + g_s^2 k_{s+1} = g_s (1 + g_s) k_{s+1} \neq 0.$$

Mit den Spalten von P_r sind auch die Spalten von P orthogonal.

- (2) P und damit T_i hat vollen Spaltenrang $(m_i - m_{i+1})$. Daß der Rang von $(\Delta_i - \Delta_{i+1})$ gleichfalls $(m_i - m_{i+1})$ ist, erkennt man aus der Zerlegung

$$I = (I - \Delta_i) + (\Delta_i - \Delta_{i+1}) + \Delta_{i+1}$$

in der

$$(I - \Delta_i)(\Delta_i - \Delta_{i+1}) = (I - \Delta_i)\Delta_{i+1} = (\Delta_i - \Delta_{i+1})\Delta_{i+1} = 0$$

und

$$\text{rg}(I - \Delta_i) = N - m_i$$

sowie

$$\text{rg}(\Delta_{i+1}) = m_{i+1}.$$

- (3) Es ist

$$\Delta_i T_i = G_i (G_i' G_i)^{-1} G_i' P (P' P)^{-1/2}$$

Jedes Element von $G_i' P$ ist die Summe aller Elemente einer Spalte von P , die einem bestimmten Aggregat in $\mathcal{Z}_i$ angehören. Also ist $(G_i' G_i)^{-1} G_i' P$ die direkte Summe von Matrizen der Form

$$= \begin{bmatrix} 1/k_1 & & & \\ & 1/k_2 & & \\ & & \ddots & \\ \emptyset & & & 1/k_p \end{bmatrix} \begin{bmatrix} k_1 & k_1 & \dots & k_1 \\ -g_1 k_2 & k_2 & \dots & k_2 \\ 0 & -g_2 k_3 & \dots & k_3 \\ \vdots & & & \\ 0 & 0 & \dots & -g_{p-1} k_p \end{bmatrix}$$

$$= \begin{bmatrix} 1 & 1 & \dots & 1 \\ -g_1 & 1 & \dots & 1 \\ 0 & -g_2 & \dots & 1 \\ \vdots & & & \\ 0 & 0 & \dots & -g_{p-1} \end{bmatrix},$$

woraus $\Delta_i T_i = T_i$ folgt.

Wegen (1) ist $G'_{i+1}P = 0$. Also

$$\Delta_{i+1}T_i = G_{i+1}(G'_{i+1}G_{i+1}P(P'P)^{-1/2}) = 0$$

und damit

$$(\Delta_i - \Delta_{i+1})T_i = T_i$$

Die gesuchte Matrix T ist also

$$T = [T_0 | T_1 | \dots | T_i | \dots | T_{n-1}]$$

und es gilt

$$T'(I - \Delta_1)T = \begin{bmatrix} I & 0 \\ (N-m_1) & \\ 0 & 0 \\ & (m_1) \end{bmatrix}$$

$$T'(\Delta_i - \Delta_{i+1})T = \begin{bmatrix} 0 & 0 & 0 \\ (n-m_i) & & \\ 0 & I & 0 \\ & (m_i - m_{i+1}) & \\ 0 & 0 & 0 \\ & & (m_{i+1}) \end{bmatrix}$$

$$T'\Delta_{n-1}T = \begin{bmatrix} 0 & 0 \\ (N-m_{n-1}) & \\ 0 & I \\ & m_{n-1} \end{bmatrix}$$

BEISPIEL: Sei $M = (1, 2, 3, \dots, 13, 14)$;

$$\mathcal{Z}_1 = (\{1, 2\}, \{3, 4, 5\}, \{6, 7\}, \{8, 9, 10\}, \{11, 12, 13, 14\}) ;$$

$$\mathcal{Z}_2 = (\{1, 2, 3, 4, 5\}, \{6, 7\}, \{8, 9, 10, 11, 12, 13, 14\}) ;$$

$$\mathcal{Z}_3 = (\{1, 2, 3, 4, 5\}, \{6, 7, 8, 9, 10, 11, 12, 13, 14\}) .$$

Dann ist T_0 vom Typ $(14, 9)$; T_1 vom Typ $(14, 2)$;

T_2 vom Typ $(14, 1)$ und T_3 vom Typ $(14, 2)$ und man findet

$$T(T'T)^{1/2} =$$

1	0	0	0	0	0	0	0	0	0	1	0	0	1	0
-1	0	0	0	0	0	0	0	0	0	1	0	0	1	0
0	1	1	0	0	0	0	0	0	0	-2/3	0	0	1	0
0	-1	1	0	0	0	0	0	0	0	-2/3	0	0	1	0
0	0	-2	0	0	0	0	0	0	0	-2/3	0	0	1	0
0	0	0	1	0	0	0	0	0	0	0	0	1	0	1
0	0	0	-1	0	0	0	0	0	0	0	0	1	0	1
0	0	0	0	1	1	0	0	0	0	0	1	-2/7	0	1
0	0	0	0	-1	1	0	0	0	0	0	1	-2/7	0	1
0	0	0	0	0	-2	0	0	0	0	0	1	-2/7	0	1
0	0	0	0	0	0	1	1	1	1	0	-3/4	-2/7	0	1
0	0	0	0	0	0	-1	1	1	1	0	-3/4	-2/7	0	1
0	0	0	0	0	0	0	-2	1	1	0	-3/4	-2/7	0	1
0	0	0	0	0	0	0	0	-3	1	0	-3/4	-2/7	0	1

Mit den Resultaten dieses Abschnitts können Determinante und Inverse von Linearkombinationen $\Sigma = c_0 I + c_1 \Delta_1 + \dots + c_{n-1} \Delta_{n-1}$ sofort explizit angegeben werden:

SATZ 1.3.4. : Sei $\Sigma = c_0 I + c_1 \Delta_1 + \dots + c_{n-1} \Delta_{n-1}$, dann gilt

- (1) $\det \Sigma = c_0^{N-m_1} (c_0 + c_1)^{m_1 - m_2} \dots (c_0 + c_1 + \dots + c_{n-1})^{m_{n-1}}$
- (2) Σ ist genau dann nichtsingulär, wenn alle Partialsummen $c_0, (c_0 + c_1), \dots, (c_0 + c_1 + \dots + c_{n-1})$ ungleich 0 sind.
- (3) Σ ist genau dann positiv definit, wenn alle Partialsummen positiv sind.
- (4) Sofern Σ nichtsingulär ist, ergibt sich

$$\Sigma^{-1} = a_0 I + a_1 \Delta_1 + \dots + a_{n-1} \Delta_{n-1}$$

$$\text{mit } a_0 := c_0^{-1}, \quad a_k = -c_k \left(\sum_{i=0}^{k-1} c_i \right)^{-1} \left(\sum_{i=0}^k c_i \right)^{-1}$$

BEWEIS

(1) folgt aus

$\det \Sigma = \det T' \Sigma T = \det \bigoplus_{i=0}^{n-1} (c_0 + c_1 + \dots + c_i) I_i$, wobei
 I_0 die Einheitsmatrix von Typ $(N-m_1, N-m_1)$, I_i
für $i = 1, 2, \dots, n-2$ die Einheitsmatrix vom Typ
 $(m_i - m_{i+1}, m_i - m_{i+1})$ und I_{n-1} die Einheitsmatrix
vom Typ (m_{n-1}, m_{n-1}) ist.

(2) und (3) sind klar.

(4) gilt wegen

$$\begin{aligned} \Sigma^{-1} &= [\lambda_0 (I - \Delta_1) + \lambda_1 (\Delta_1 - \Delta_2) + \dots + \lambda_{n-1} \Delta_{n-1}]^{-1} \\ &= \frac{1}{\lambda_0} (I - \Delta_1) + \frac{1}{\lambda_1} (\Delta_1 - \Delta_2) + \dots + \frac{1}{\lambda_{n-1}} \Delta_{n-1} \\ &= \frac{1}{c_0} (I - \Delta_1) + \frac{1}{c_0 + c_1} (\Delta_1 - \Delta_2) + \dots + \frac{1}{c_0 + c_1 + \dots + c_{n-1}} \Delta_{n-1} \\ &= \frac{1}{c_0} I + \left(\frac{1}{c_0 + c_1} - \frac{1}{c_0} \right) \Delta_1 + \dots + \left(\frac{1}{c_0 + c_1 + \dots + c_{n-1}} - \frac{1}{c_0 + c_1 + \dots + c_{n-2}} \right) \Delta_{n-1} \end{aligned}$$

1.4 KOVARIANZTHEOREME

Das "Kovarianztheorem" von ALKER [1969] wird von Sozialforschern häufig zur Interpretation mehrerebenenanalytischer Daten verwendet. Es soll hier zunächst dazu dienen, den Nutzen der eingeführten Operatorenschreibweise zu illustrieren. Der Beweis des Satzes bei ALKER (siehe auch HUMMEL [1972]) ist umständlich und zieht sich über mehrere Seiten hin - der Satz ergibt sich hier als unmittelbare Folgerung aus den elementaren Eigenschaften der Aggregationsoperatoren.

Wir betrachten zunächst den Zweiebenenfall. Eine (N, k) -Matrix X von Beobachtungen für die N Objekte in $\mathcal{U}$ auf k Variablen kann in 3 Komponenten zerlegt werden:

$$X = (I - \Delta_1)X + (\Delta_1 - M)X + MX$$

Dabei ist $M = \frac{1}{N} 1_N 1_N'$ eine idempotente (N, N) -Matrix
 $M = (m_{ij}) = (\frac{1}{N})$. M läßt sich als Aggregationsmatrix der größtmöglichen Zerlegung von $\mathcal{U}$ interpretieren, in der alle Elemente von $\mathcal{U}$ in einem Aggregat zusammengefaßt werden. Diese Zerlegung ist eine Vergrößerung aller Zerlegungen der Mehrebenen-situation, es gilt folglich mit Satz 1.2.2., daß $\Delta_1 M = M \Delta_1 = M$ für $i = 1 \dots n-1$. Die drei Komponenten in der Aufspaltung von X sind daher orthogonal (Satz 1.3.1.) und können inhaltlich als

- Abweichungen der individuellen Werte von den Gruppenmitteln $((I - \Delta_1)X)$
- Abweichungen der Gruppenmittel vom Gesamtmittel $((\Delta_1 - M)X)$
- Gesamtmittelwerte (MX)

interpretiert werden.

Die beobachtete Varianz-Kovarianzmatrix der Spalten von X ist

$$C_X = \frac{1}{N} X' (I - M) X$$

und das Kovarianztheorem zeigt, daß C_X die Summe zweier Matrizen $C_{(I - \Delta_1)X}$ und $C_{\Delta_1 X}$ ist, die sich ebenfalls als Kovarianzmatrizen interpretieren lassen.

$C_{(I - \Delta_1)X}$ ist die Kovarianzmatrix der individuellen Abweichungen von den Gruppenmitteln (Kovarianz "innerhalb" der Aggregate, Mikrokovarianz).

$C_{\Delta_1 X}$ ist die Kovarianzmatrix der aggregierten Werte (Kovarianz "zwischen" den Aggregaten, Makrokovarianz).

SATZ 1.4.1. : Kovarianztheorem (Zweiebenenfall):

$$C_X = C_{(I - \Delta_1)X} + C_{\Delta_1 X}, \text{ wobei}$$

$$C_{(I - \Delta_1)X} := \frac{1}{N} X' (I - \Delta_1) (I - M) (I - \Delta_1) X = \frac{1}{N} X' (I - \Delta_1) X$$

$$C_{\Delta_1 X} := \frac{1}{N} X' \Delta_1 (I - M) \Delta_1 X = \frac{1}{N} X' (\Delta_1 - M) X$$

BEWEIS : Wir gehen in der Zerlegung

$$X = (I - \Delta_1)X + (\Delta_1 - M)X + MX$$

zu Abweichungen von den Gesamtmitteln über:

$$(I - M)X = (I - M)(I - \Delta_1)X + (I - M)(\Delta_1 - M)X .$$

Die Komponenten sind orthogonal und es gilt

$$(I - M)(I - \Delta_1) = I - M - \Delta_1 + M = (I - \Delta_1)$$

$$(I - M)(\Delta_1 - M) = (I - M)\Delta_1 = (\Delta_1 - M) ,$$

so daß

$$X'(I - M)X = X'(I - \Delta_1)X + X'(\Delta_1 - M)X .$$

Das Kovarianztheorem wird in der mehrbenenanalytischen Literatur gerne zum Ausgangspunkt einer Diskussion sogenannter "ökologischer Fehlschlüsse" gewählt. Dabei geht es um die Möglichkeit, Ergebnisse, die auf einer Datenebene gewonnen wurden, auf andere Datenebenen zu übertragen - konkret etwa darum, ob es möglich ist, aus dem positiven Vorzeichen einer Makrokovarianz auch auf eine positive Mikro- oder Gesamtkovarianz zu schließen.

Derartige Diskussionen sind wenig aussagekräftig, solange sie rein auf der deskriptiven Ebene, d.h. ohne die Hilfe stochastischer Modelle geführt werden. Nimmt man etwa an, daß die N Zeilen von X unabhängige Realisierungen eines Zufallsvektors mit Erwartungswert μ und Kovarianzmatrix Ω darstellen, so wäre der oben genannte Schluß von der "Makro"- auf die Gesamtkovarianz durchaus korrekt: man findet dann nämlich für die Erwartungswerte der Kovarianzmatrix (siehe Kapitel 3)

$$E[C_X] = \frac{N-1}{N} \Omega$$

$$E[C_{\Delta X}] = \frac{m-1}{N} \Omega$$

$$E[C_{(I-\Delta)X}] = \frac{N-m}{N} \Omega .$$

Jede der drei Kovarianzmatrizen kann also zur unverzerrten Schätzung für das gleiche Ω benutzt werden - da auch die Varianzen der Schätzer mit wachsendem $(N-m)$ und m abnehmen, darf man in "großen Stichproben" erwarten, daß

$\frac{N}{N-1} C_X$, $\frac{m-1}{N} C_{\Delta X}$ und $\frac{N}{N-m} C_{(I-\Delta)X}$ nur wenig voneinander abweichen.

Im n-Ebenenfall geht man analog von einer Zerlegung von X in $n+1$ orthogonale Komponenten aus:

$$X = (I-\Delta_1)X + (\Delta_1-\Delta_2)X + \dots + (\Delta_{n-1}-M)X + MX.$$

Diese Komponenten sind inhaltlich als Gesamtmittel bzw. Abweichungen zwischen je zwei aufeinanderfolgenden Ebenen zu interpretieren. Das Kovarianztheorem zerlegt die Gesamtkovarianzmatrix C_X in n Kovarianzmatrizen, die diesen Abweichungen entsprechen.

SATZ 1.4.2. : Kovarianztheorem (n-Ebenenfall)

$$C_X = C_{(I-\Delta_1)X} + \dots + C_{(\Delta_{n-1}-M)X}$$

BEWEIS : Analog zu 1.4.1.

2. LINEARE MEHREBENENMODELLE

2.0. VORBEMERKUNG

In linearen Modellen wird üblicherweise angenommen, daß der Vektor der Beobachtungen $y = (y_1, y_2, \dots, y_N)'$ einer Verteilung mit Erwartungswert $E[y] = X\beta$ und Varianz-Kovarianzmatrix $\text{var}[y] = \sigma^2 I$ entstammt. (Anm. 1). Dabei ist X eine nicht-stochastische (N, k) -Matrix mit vollem Spaltenrang $k < N$, β und σ^2 sind feste, aber unbekannte Parameter. Trifft man die zusätzliche Annahme, daß die Verteilung von y eine multivariate Normalverteilung ist, so kann das Modell koncis: $y \sim N[X\beta, \sigma^2 I]$ geschrieben werden.

Unsere Kritik richtete sich gegen die Verwendung der Annahme $\text{var}[y] = \sigma^2 I$ in Mehrebenen-situationen. Im ersten Abschnitt werden Annahmen über die Kovarianzmatrix von y vorgeschlagen, die (hoffentlich) substanzwissenschaftlich plausibler sind als die konventionelle Annahme unkorrelierter Beobachtungen. Das entstehende Modell ist eine Verallgemeinerung des Intraklassen-korrelationsmodells von HALPERIN [1951] auf mehr als zwei Ebenen und auf Gruppen unterschiedlicher Größe.

Anm. 1: Bei der Behandlung linearer Modelle widersprechen sich zwei Konventionen der Schreibweise - in der linearen Algebra werden Vektoren üblicherweise klein, Matrizen groß geschrieben; in der Wahrscheinlichkeitstheorie werden Zufallsvariable groß, feststehende Größen klein geschrieben. Wir entscheiden uns hier für die Konventionen der linearen Algebra - unterscheiden also in der Notation nicht zwischen Zufallsvariablen und ihre Realisierungen.

Eine Reihe ganz unterschiedlicher Ansätze führen zum gleichen Modell; mögliche Interpretationen der Modellparameter, die sich aus diesen Ansätzen ergeben, werden im zweiten Abschnitt diskutiert.

Da bisher in Mehrebenensituationen ausschließlich OLS-Schätzer $\beta^* = (X'X)^{-1}X'y$ für β benutzt wurden, untersucht der dritte Abschnitt das Verhalten dieser Schätzer in Mehrebenenmodellen der vorgeschlagenen Kovarianzstruktur.

2.1. EINFÜHRUNG DES MODELLS

Wir beginnen die Diskussion möglicher Alternativen zur Annahme $\text{var}[y] = \sigma^2 I$ mit dem überschaubareren Zweiebenenfall.

Gegeben ist also das Paar $(\mathcal{M}, \mathcal{Z})$, in dem $\mathcal{Z}$ eine Zerlegung von $\mathcal{M}$ in m Aggregate $\mathcal{M}_1, \mathcal{M}_2, \dots, \mathcal{M}_m$ darstellt, und eine Realisierung des N -dimensionalen Zufallsvektors y aus einer Verteilung mit Erwartungswertvektor $X\beta$. Dabei ist X eine nichtstochastische Matrix vom Typ (N, k) mit vollem Spaltenrang k .

Für Beobachtungen y_s und y_t aus verschiedenen Aggregaten wird wie bisher $\text{cov}[y_s, y_t] = 0$ angenommen. Seien also y_s und y_t Beobachtungen bei Elementen des gleichen Aggregats $\mathcal{M}_j$ aus $\mathcal{Z}$. Zwei zunächst plausible alternative Annahmen über die Kovarianz von y_s und y_t (Intragruppenkovarianz) sind:

- (1) $\text{cov}[y_s, y_t] = \alpha_j$ (möglicherweise verschiedene Intragruppenkovarianzen für die m Gruppen)
- (2) $\text{cov}[y_s, y_t] = \alpha$ (konstante Intragruppenkovarianzen).

Beide Annahmen verlieren bei näherer Betrachtung viel von ihrer Plausibilität:

Annahme (1) ist vom statistischen Standpunkt her bedenklich. In

typischen Anwendungen der Mehrebenenanalyse (insbesondere in der Organisationsforschung) kann die Zahl der zu schätzenden Parameter α_j in der Größenordnung einiger hundert liegen - wie wir im letzten Kapitel zeigen werden, beruhen Schätzer für α_j nur auf einem Freiheitsgrad und sind dementsprechend sehr instabil.

Annahme (2) ist aus soziologischen Gründen wenig attraktiv, sobald die auftretenden Aggregate unterschiedlich groß sind. Wesentliche Faktoren für die Existenz von Abhängigkeiten zwischen den Mitgliedern eines Aggregats wie Interaktionsdichte, Homogenität der Gruppe, Gleichartigkeit externer Anforderungen an die Gruppenmitglieder variieren nämlich mit dem Umfang der Gruppe: mit steigender Gruppengröße wird also die Intragruppenkovarianz nicht konstant bleiben, sondern abnehmen.

Die spezifische Annahme, daß Intragruppenkovarianzen invers proportional zur Gruppengröße variieren, also daß

$$(3) \text{ cov}[y_s, y_t] = \frac{\alpha}{n_j} \quad (\text{wobei } n_j \text{ die Anzahl der Elemente in } \mathcal{U}_j \text{ ist}) ,$$

führt zu einem Modell mit besonders angenehmen mathematischen Eigenschaften und ergibt sich auch, wenn man einen der alternativen Ansätze des nächsten Abschnittes zum Ausgangspunkt wählt.

Unter Hinzunahme eines Terms für die Varianz der Beobachtungen y_s findet man im Zweiebenenfall

$$\begin{aligned} \text{var}[y] &= \bigoplus_{j=1}^m \alpha_0 I_{n_j} + \frac{\alpha}{n_j} \mathbb{1}_{n_j} \mathbb{1}_{n_j} \\ &= \alpha_0 I + \alpha \Delta \end{aligned}$$

Analoge Überlegungen für den n-Ebenenfall führen auf

$$\text{var}[y] = \alpha_0 I + \alpha_1 \Delta_1 + \dots + \alpha_{n-1} \Delta_{n-1} ,$$

dabei ist also die Intragruppenkovarianz eines Aggregats β_{ir}

der Zerlegung β_i mit n_{ir} Elementen durch $\frac{\alpha_i}{n_{ir}}$ gegeben.

Die Ergebnisse von Kapitel 1 erlauben eine Reparametrisierung von $\text{var}[y]$, die das Schätzproblem für die Parameter der Kovarianzmatrix wesentlich vereinfacht. Diese Reparametrisierung besteht im Übergang zu den Eigenwerten λ_i der Kovarianzmatrix und den zugehörigen Hauptidepotenten K_i :

$$\text{var}[y] = \lambda_0 K_0 + \lambda_1 K_1 + \dots + \lambda_{n-1} K_{n-1} \quad ,$$

wobei $K_0 = I - \Delta_1$, $K_i = \Delta_i - \Delta_{i+1}$ ($i=1 \dots n-2$), $K_{n-1} = \Delta_{n-1}$ und $\lambda_i = \alpha_0 + \alpha_1 + \dots + \alpha_i$ ($i = 0 \dots n-1$).

Es ist möglich, eine analoge Kovarianzstruktur für das multivariate Regressionsmodell zu definieren.

Gegeben ist hier eine (N,p) -Matrix $Y = [y_1 | y_2 | \dots | y_p]$, die eine Realisierung einer zufälligen Matrix mit dem Erwartungswert $E[Y] = XB$ darstellt. Dabei ist X wieder eine nichtstochastische (N,k) -Matrix mit vollem Spaltenrang und B eine (k,p) -Matrix fester, aber unbekannter Parameter. Für die Kovarianzmatrix der Spalten y_i und y_j von Y :

$$\text{cov}[y_i, y_j] = E[(y_i - E[y_i])(y_j - E[y_j])']$$

wird eine zu $\text{var}[y]$ analoge Kovarianzstruktur angenommen:

$$\text{cov}[y_i, y_j] = w_{ij}^{(0)} K_0 + w_{ij}^{(1)} K_1 + \dots + w_{ij}^{(n-1)} K_{n-1} \quad .$$

Schreibt man $\text{vec } Y = [y_{11} y_{12} \dots y_{1p} | y_{21} y_{22} \dots y_{2p} | \dots | y_{N1} \dots y_{Np}]'$ so ist also

$$\text{var}[\text{vec } Y] = K_0 \otimes \Omega_0 + K_1 \otimes \Omega_1 + \dots + K_{n-1} \otimes \Omega_{n-1}$$

wobei $\Omega_s = (w_{ij}^{(s)})$, $s = 0, 1 \dots n-1$ und $A \otimes B$ das Kroneckerprodukt der Matrizen A und B bedeutet (siehe etwa BELLMAN [1960]).

BEISPIEL: In einer Zweiebenensituation $(U, \mathcal{Z})$ mit $U = (1, 2, 3, 4, 5)$ und $\mathcal{Z} = (\{1, 2, 3\}, \{4, 5\})$ hat die Kovarianzmatrix von $(\text{vec } Y)$ folgende Form:

$$\text{var}[\text{vec } Y] = \begin{bmatrix} 2/3\Omega_0 & -1/3\Omega_0 & -1/3\Omega_0 & 0 & 0 \\ -1/3\Omega_0 & 2/3\Omega_0 & -1/3\Omega_0 & 0 & 0 \\ -1/3\Omega_0 & -1/3\Omega_0 & 2/3\Omega_0 & 0 & 0 \\ 0 & 0 & 0 & 1/2\Omega_0 & -1/2\Omega_0 \\ 0 & 0 & 0 & -1/2\Omega_0 & 1/2\Omega_0 \end{bmatrix} + \begin{bmatrix} 1/3\Omega_1 & 1/3\Omega_1 & 1/3\Omega_1 & 0 & 0 \\ 1/3\Omega_1 & 1/3\Omega_1 & 1/3\Omega_1 & 0 & 0 \\ 1/3\Omega_1 & 1/3\Omega_1 & 1/3\Omega_1 & 0 & 0 \\ 0 & 0 & 0 & 1/2\Omega_1 & 1/2\Omega_1 \\ 0 & 0 & 0 & 1/2\Omega_1 & 1/2\Omega_1 \end{bmatrix}$$

Wir beschränken uns in dieser Arbeit jedoch auf das univariate Modell. Ein Spezialfall des Modells ist in der statistischen Literatur als "Intraklassenkorrelationsmodell" von HALPERIN [1951] bekannt. Dieser Spezialfall ergibt sich, wenn wir eine Zweiebenensituation betrachten, in der alle Aggregate gleich groß sind.

In der Formulierung von HALPERIN lautet dann die Varianz-Kovarianzmatrix von y :

formidable character that an explicit solution does not appear possible".

2.2. INTERPRETATIONEN DER PARAMETER

Das vorgeschlagene Modell kann mit einer Reihe von unterschiedlichen Ansätzen gewonnen werden, die Interpretationsmöglichkeiten für verschiedene Funktionen der Parameter des Modells liefern:

- (1) direkt aus Annahmen über die Abhängigkeit der Intragruppenkovarianzen von der Gruppengröße.
- (2) als Varianzkomponentenmodell, wenn man die Spektraldarstellung von $\text{var}[y]$ zugrundelegt.
- (3) als Modell, in dem der Wert eines Individuums auf der abhängigen Variablen nicht nur von einer individuellen Störung, sondern auch vom Durchschnitt der Störungen in der jeweiligen Gruppe abhängt.

Bei der Darstellung der Ansätze beschränken wir uns auf den Zweiebenenfall, da sich beim Übergang zu n Ebenen nichts Grundlegendes ändert. Der Ansatz (1) wurde hier zur Einführung des Modells gewählt. Die Intragruppenkovarianz einer Gruppe 3_{1j} mit r Mitgliedern ist (im Zweiebenenfall)

$$\alpha_j = \frac{\lambda_1 - \lambda_0}{r}$$

Da Sozialwissenschaftler gewöhnt sind, mit Korrelationskoeffizienten umzugehen, seien die entsprechenden "Intragruppenkorrelationen" ρ_j angegeben. Die Varianz der Störungen in einer

Gruppe mit r Mitgliedern ist $\lambda_0 + \frac{1}{r}(\lambda_1 - \lambda_0)$, man findet

also

$$\rho_j = \frac{\lambda_1 - \lambda_0}{\lambda_1 + (r-1)\lambda_0}$$

Zur Interpretation sind diese Größen wegen der Abhängigkeit von r nur bedingt geeignet - zwei Spezialfälle verdienen allerdings Beachtung:

Für die kleinstmögliche Gruppe mit $r = 2$ Mitgliedern ist der Intragruppenkorrelationskoeffizient

$$p = \frac{\lambda_1 - \lambda_0}{\lambda_1 + \lambda_0}$$

Für eine Gruppe mit durchschnittlicher Gruppengröße ($r = \frac{N}{m}$) ist

$$p = \frac{\lambda_1 - \lambda_0}{\lambda_1 + \frac{N}{m} \lambda_0} .$$

Ansatz (2) geht direkt von den Eigenwerten λ_0, λ_1 der Varianz-Kovarianzmatrix aus (bzw. im multivariaten Modell von den Matrizen Ω_0 und Ω_1). Diese Größen entsprechen auf der theoretischen Ebene den Komponenten in der Zerlegung der beobachteten Varianz (bzw. Varianz-Kovarianzmatrix) in "Mikro"- und "Makro"bestandteile. Der Erwartungswert der Mikrovarianz der Störungen $u = y - X\beta$ ist nämlich im vorgeschlagenen Modell gleich (siehe Kapitel 3)

$$E\left[\frac{1}{N} u'(I - \Delta)u\right] = \frac{N-m}{N} \lambda_0 ,$$

der Erwartungswert der Makrovarianz ist

$$E\left[\frac{1}{N} u'(\Delta - M)u\right] = \frac{m-1}{N} \lambda_1 .$$

Im multivariaten Modell findet man für die Erwartungswerte entsprechend

$$E\left[\frac{1}{N} U'(I - \Delta)U\right] = \frac{N-m}{N} \Omega_0$$

$$E\left[\frac{1}{N} U'(\Delta - M)U\right] = \frac{m-1}{N} \Omega_1$$

- dabei ist $U = Y - XB$ die Matrix der Störungen des Modells.

Die Modellgleichung im Ansatz (3) lautet:

$$y = XB + (\sigma_0 I + \sigma_1 \Delta) e .$$

Dabei ist e ein Vektor mit individuellen Störungen, für den $E[e] = 0$ und $\text{var}[e] = I$ angenommen wird - die Parameter σ_0 und σ_1 beschreiben, wie stark die individuellen Störungen (e) bzw. die Gruppendurchschnitte der Störungen (Δe) zur Bestimmung von y beitragen. (Im Modell mit unkorrelierten Beobachtungen ist $\sigma_1 = 0$ und $\sigma_0^2 = \sigma^2$, also gleich der Varianz einer Beobachtung.)

Für die Varianz-Kovarianzmatrix des zufälligen Teils dieses Modells findet man

$$\begin{aligned} \text{var}[(\sigma_0 I + \sigma_1 \Delta) e] &= \sigma_0^2 I + (\sigma_1^2 + 2\sigma_0 \sigma_1) \Delta \\ &= \sigma_0^2 (I - \Delta) + (\sigma_0 + \sigma_1)^2 \Delta , \end{aligned}$$

die Parameter σ_0 und σ_1 lassen sich also in λ_0 und λ_1 wie folgt ausdrücken:

$$\begin{aligned} \sigma_0 &= \sqrt{\lambda_0} \\ \sigma_1 &= \sqrt{\lambda_1} - \sqrt{\lambda_0} . \end{aligned}$$

Ansatz (3) liefert ein zusätzliches heuristisches Argument für die Einführung von Intragruppenkovarianzen, die nicht konstant für alle Gruppen, sondern invers proportional zur Gruppengröße sind: in diesem Ansatz würden konstante Intragruppenkovarianzen nämlich bedeuten, daß eine Abhängigkeit der zu erklärenden Variable nicht von den Mittelwerten, sondern von den Summen der Störungen in jeder Gruppe postuliert würde.

2.3. OLS-SCHÄTZER IN MEHREBENENMODELLEN

Die Anwendung der OLS-Schätzfunktion $\beta^* = (X'X)^{-1}X'y$ in Mehrebenenmodellen mit korrelierten Beobachtungen stößt auf zwei Probleme:

- (1) β^* ist im allgemeinen eine ineffiziente Schätzfunktion für β
- (2) die übliche Schätzung für die Varianz-Kovarianzmatrix von β^* , also $\hat{\sigma}^2(X'X)^{-1}$, wobei

$$\hat{\sigma}^2 = (N-k)^{-1}(y - X\beta^*)'(y - X\beta^*),$$

kann ein ungeeigneter Schätzer für die tatsächliche Varianz-Kovarianzmatrix sein.

Wir zeigen im folgenden, daß diese beiden Probleme von unterschiedlicher Größenordnung sind - zur Illustration diene zunächst der Fall eines Zweiebenenmodells mit einem Regressor, also das Modell

$$y = x\beta + u \quad \text{mit} \quad \text{var}[u] = \lambda_0 K_0 + \lambda_1 K_1.$$

2.3.1. SPEZIALFALL: ZWIEEBENENMODELL MIT EINEM REGRESSOR

Relative Effizienz von β^* : Wir vergleichen β^* mit der GAUSS-MARKOV-Schätzfunktion

$$\hat{\beta} = \frac{\frac{1}{\lambda_0} x'K_0 y + \frac{1}{\lambda_1} x'K_1 y}{\frac{1}{\lambda_0} x'K_0 x + \frac{1}{\lambda_1} x'K_1 x}$$

Unter Normalverteilungsannahmen hat $\hat{\beta}$ minimale Varianz unter

allen unverzerrten Schätzfunktionen für β .

Als relative Effizienz von β^* gegenüber β definieren wir

$$\begin{aligned} \text{eff} &:= \frac{\text{var}(\hat{\beta})}{\text{var}(\beta^*)} \\ &= \frac{\left(\frac{1}{\lambda_0} x'K_0x + \frac{1}{\lambda_1} x'K_1x\right)^{-1}}{(x'x)^{-1}(\lambda_0 x'K_0x + \lambda_1 x'K_1x)(x'x)^{-1}} \end{aligned}$$

Umformungen ergeben:

$$\text{eff} = \frac{1}{1 + \frac{(\lambda_0 - \lambda_1)^2}{\lambda_0 \lambda_1} \left(1 - \frac{x' \Delta x}{x'x}\right) \left(\frac{x' \Delta x}{x'x}\right)}$$

Die Ineffizienz von β^* wird also von zwei Faktoren bestimmt:

- dem Ausdruck $\frac{(\lambda_0 - \lambda_1)^2}{\lambda_0 \lambda_1}$, der mit wachsender Korrelation der Beobachtungen ansteigt
- dem Ausdruck $\left(1 - \frac{x' \Delta x}{x'x}\right) \left(\frac{x' \Delta x}{x'x}\right)$, der angibt, wie weit sich x von den Extremen einer reinen Makrovariable ($x'x = x' \Delta x$) bzw. einer reinen Mikrovariable ($x' \Delta x = 0$) entfernt.

Eine untere Schranke für eff ist durch die Beziehung

$$\text{eff} \geq \frac{1}{1 + \frac{1}{4} \frac{(\lambda_0 - \lambda_1)^2}{\lambda_0 \lambda_1}} = \frac{4}{\alpha + \frac{1}{\alpha} + 2}$$

gegeben, wobei $\alpha = \frac{\lambda_1}{\lambda_0}$ gesetzt wurde. Die folgende Tabelle

zeigt, daß erst bei relativ unrealistischen Werten für das Verhältnis $\frac{\lambda_1}{\lambda_0}$ die Effizienzverluste für β^* bedenklich werden können:

λ_1/λ_0	untere Schranke für eff
1.0	1.00
1.5	0.96
2.0	0.89
2.5	0.82
3.0	0.75
3.5	0.69
4.0	0.64
4.5	0.60
5.0	0.56

Fehleinschätzung der Variabilität von β^*

Die konventionelle Schätzfunktion für $\text{var}[\beta^*]$ ist

$$\begin{aligned} v_* &= \hat{\sigma}^2 (x'x)^{-1} \\ &= \frac{1}{N-1} \frac{y'(I - x(x'x)^{-1}x')y}{x'x} \end{aligned}$$

Unterstellen wir das Modell

$$\text{var}[y] = \lambda_0 K_0 + \lambda_1 K_1 ,$$

so ist die wahre Varianz von β^*

$$v = (x'x)^{-1} (\lambda_0 x'K_0x + \lambda_1 x'K_1x) (x'x)^{-1} .$$

Es liegt nahe, den Ausdruck

$$T := \frac{E[v_*]}{v}$$

zu untersuchen; man findet

$$T = \frac{1}{N-1} \frac{\lambda_0 (N-m) + \lambda_1 m}{\lambda_0 \left(\frac{x'K_0x}{x'x} \right) + \lambda_1 \left(\frac{x'K_1x}{x'x} \right)} - \frac{1}{N-1} .$$

Setzen wir $\alpha := \frac{\lambda_1}{\lambda_0}$, $w_1 = \frac{x'K_1x}{x'x}$, so ist

$$T = \frac{1 + (\alpha-1)\left(\frac{m-w_1}{N-1}\right)}{1 + (\alpha-1)w_1}$$

Für $\alpha > 1$ (positive Intragruppenkorrelation) ist dies eine monoton fallende Funktion von w_1 - Unterschätzung von $\text{var}[\beta^*]$ tritt ein für

$$\frac{m-w_1}{N-1} < w_1$$

bzw.

$$w_1 > \frac{m}{N}$$

Maximale Unterschätzung findet man für $w_1 = 1$ - es gilt also

$$T \geq \frac{1}{\alpha} + (1 - \frac{1}{\alpha})\left(\frac{m-1}{N-1}\right) = \frac{1}{\alpha} \left(1 - \frac{m-1}{N-1}\right) + \frac{m-1}{N-1}$$

Maximale Überschätzung findet man für $w_1 = 0$:

$$T \leq 1 + (\alpha-1) \frac{m}{N-1} = \alpha \left(\frac{m}{N-1}\right) + \left(\frac{N-m-1}{N-1}\right)$$

Ober- und Untergrenzen für T zeigt die folgende Tabelle:

α	Obergrenze für T	Untergrenze für T
1	1	1
2	$1 + \frac{m}{N-1}$	$\frac{1}{2} + \frac{1}{2} \left(\frac{m-1}{N-1}\right)$
3	$1 + 2 \frac{m}{N-1}$	$\frac{1}{3} + \frac{2}{3} \left(\frac{m-1}{N-1}\right)$
4	$1 + 3 \frac{m}{N-1}$	$\frac{1}{4} + \frac{3}{4} \left(\frac{m-1}{N-1}\right)$
5	$1 + 4 \frac{m}{N-1}$	$\frac{1}{5} + \frac{4}{5} \left(\frac{m-1}{N-1}\right)$

Da die Anzahl der Gruppen (m) im allgemeinen klein im Verhältnis zur Anzahl der beobachteten Individuen (N) sein wird, kann die Unterschätzung von $\text{var}[\beta^*]$ insbesondere für solche Regressoren bedrohlich werden, die innerhalb der Gruppen konstant oder nahezu konstant sind ($w_1 \rightarrow 1$).

2.3.2. ALLGEMEINER FALL

Im allgemeinen Fall (k Regressoren und n Ebenen) betrachten wir zur Beurteilung der Effizienz von $\beta^* = (X'X)^{-1}X'y$ das Verhältnis der generalisierten Varianz von β^* und $\hat{\beta} = (X'\Sigma^{-1}X)^{-1}X'\Sigma^{-1}y$, also

$$\begin{aligned} \text{eff} &= \frac{\det \text{var}[\hat{\beta}]}{\det \text{var}[\beta^*]} \\ &= \frac{\det[X'\Sigma^{-1}X]^{-1}}{\det[X'X]^{-1}X'\Sigma X[X'X]^{-1}} \end{aligned}$$

Dabei ist $\Sigma = \lambda_0 K_0 + \lambda_1 K_1 + \dots + \lambda_{n-1} K_{n-1}$.

In Mehrebenensituationen wird man im allgemeinen positive Korrelationen der Beobachtungen annehmen können, also

$\lambda_{n-1} \geq \lambda_{n-2} \geq \dots \geq \lambda_1 \geq \lambda_0$. Wir bezeichnen das Verhältnis des größten Eigenwerts von Σ zum kleinsten mit α (in der Regel ist also $\alpha = \lambda_{n-1} \lambda_0^{-1}$).

SATZ 2.3.2.1. : Es gilt

$$1 \geq \text{eff} \geq \left[\frac{4}{\alpha + \frac{1}{\alpha} + 2} \right]^k$$

BEWEIS : Die linke Seite der Ungleichung ist klar, da $\text{var}[\beta^*] - \text{var}[\hat{\beta}]$ positiv definit ist (GAUSS-MARKOFF-

Theorem). Da X spaltenregulär ist, existiert eine orthogonale (k,k) -Matrix P und eine Diagonalmatrix D mit positiven Diagonalelementen, so daß

$$X'X = PDP'$$

oder

$$I = D^{-1/2}P'X'XPD^{-1/2}.$$

Wir definieren eine (N,k) -Matrix Y durch

$$Y = XPD^{-1/2}.$$

Die Spalten $[y_1 | y_2 | \dots | y_k]$ von Y sind orthonormiert, nach Definition gilt

$$Y'Y = I.$$

Mit $X = YD^{1/2}P'$ findet man also

$$\begin{aligned} \frac{1}{\text{eff}} &= \frac{\det(X'EX) \det(X'\Sigma^{-1}X)}{\det(X'X) \det(X'X)} \\ &= \frac{\det(PD^{1/2}Y'\Sigma YD^{1/2}P') \det(PD^{1/2}Y'\Sigma^{-1}YD^{1/2}P')}{\det(PD^{1/2}Y'YD^{1/2}P') \det(PD^{1/2}Y'YD^{1/2}P')} \\ &= \det(Y'\Sigma Y) \det(Y'\Sigma^{-1}Y). \end{aligned}$$

Nun gilt für jede positiv-definite Matrix $A = (a_{ij})_{(n,n)}$, daß

$$\det(A) \leq a_{11}a_{22}\dots a_{nn}$$

folglich

$$\frac{1}{\text{eff}} \leq (y_1'\Sigma y_1)(y_2'\Sigma y_2)\dots(y_k'\Sigma y_k)(y_1'\Sigma^{-1}y_1)(y_2'\Sigma^{-1}y_2)\dots(y_k'\Sigma^{-1}y_k)$$

Wir benutzen die Ungleichung von KANTOROWICH, nach der für

jeden Einheitsvektor z und jede positiv-definite, symmetrische Matrix A

$$(z'Az)(z'A^{-1}z) \leq \frac{1}{4} \left(\alpha + \frac{1}{\alpha} + 2 \right)^k,$$

wobei α das Verhältnis des größten Eigenwerts zum kleinsten Eigenwert von A ist.

Hier ist also

$$\frac{1}{\text{eff}} \leq \left(\frac{\alpha + \frac{1}{\alpha} + 2}{4} \right)^k,$$

woraus die Behauptung folgt.

Die gewonnene Abschätzung für die Effizienz von β^* ist analog zu der im Spezialfall $k = 1$ erhaltenen Schranke und führt zu analogen Schlußfolgerungen.

Um die Fehleinschätzung von $\text{var}[\beta^*]$ durch Anwendung von

$$\begin{aligned} V_* &= \hat{\sigma}^2 (X'X)^{-1} \\ &= \frac{1}{N-k} [y'(I - X(X'X)^{-1}X')y] (X'X)^{-1} \end{aligned}$$

anstelle von

$$V = (X'X)^{-1} (X'\Sigma X) (X'X)^{-1}$$

beurteilen zu können, betrachten wir den Ausdruck

$$T = \frac{\det E[V_*]}{\det V}.$$

Man findet für $E[V_*]$:

$$E[V_*] = \frac{\text{tr} E - \text{tr}(X'X)^{-1}X'\Sigma X}{N - k} (X'X)^{-1}$$

und folglich

$$T = \frac{\text{tr} E - \text{tr}(X'X)^{-1}X'\Sigma X}{N - k} \cdot \frac{\det(X'X)}{\det(X'\Sigma X)}.$$

Nehmen wir an, daß $\lambda_{n-1} \geq \lambda_{n-2} \geq \dots \geq \lambda_1 \geq \lambda_0$ und bezeichnen wir wie oben das Verhältnis λ_{n-1}/λ_0 mit α , so liefert der folgende Satz eine Abschätzung für T in α bzw. $\frac{1}{\alpha}$.

SATZ 2.3.2.2. : Es gilt

$$\left[\alpha \left(\frac{m_1}{N-k} \right) + \left(\frac{N-m_1-k}{N-k} \right) \right]^k \geq T \geq \left[\frac{1}{\alpha} \left(\frac{N-m_{n-1}}{N-k} \right) + \left(\frac{m_{n-1}-k}{N-k} \right) \right]^k$$

BEWEIS : Wie in Satz 2.3.2.1. gehen wir zu einer Orthonormalbasis Y des Spaltenraums von X über und erhalten

$$T = \left[\frac{\text{tr} E - \text{tr} Y' E Y}{N-k} \right]^k \frac{1}{\det Y' E Y}.$$

Nun gilt

$$\det Y' E Y \leq \lambda_{n-1}^k$$

also

$$\begin{aligned} T &\geq \left[\frac{\text{tr} E - \text{tr} Y' E Y}{N-k} \right]^k \frac{1}{\lambda_{n-1}^k} \\ &= \left[\frac{1}{N-k} \sum_{i=0}^{n-1} \frac{\lambda_i}{\lambda_{n-1}} (\text{tr} K_i) - \frac{1}{N-k} \sum_{i=0}^{n-1} \frac{\lambda_i}{\lambda_{n-1}} (\text{tr} Y' K_i Y) \right]^k \\ &\geq \left[\frac{1}{N-k} \sum_{i=0}^{n-2} \frac{1}{\alpha} (\text{tr} K_i) - \frac{1}{N-k} \sum_{i=0}^{n-2} \frac{1}{\alpha} (\text{tr} Y' K_i Y) \right. \\ &\quad \left. + \frac{1}{N-k} \text{tr} K_{n-1} - \frac{1}{N-k} \text{tr} Y' K_{n-1} Y \right]^k \\ &\geq \left[\frac{1}{N-k} \frac{1}{\alpha} (n-m_{n-1}) + \frac{m_{n-1}}{N-k} - \frac{k}{N-k} \right]^k \end{aligned}$$

$$\text{da } \sum_{i=0}^{n-1} \text{tr} Y' K_i Y = k \quad \text{und} \quad \frac{1}{\alpha} < 1.$$

Es folgt

$$T \geq \left[\frac{1}{\alpha} \left[\frac{N-m}{N-k} \right] + \frac{m}{N-k} \right]^k .$$

Da die Spalten von Y orthogonal sind, gilt

$$\det Y' \Sigma Y \geq \lambda_0^k$$

womit

$$\begin{aligned} T &\leq \left(\frac{\text{tr} \Sigma - \text{tr} Y' \Sigma Y}{N-k} \right)^k \frac{1}{\lambda_0^k} \\ &= \left[\frac{1}{N-k} \sum_{i=0}^{n-1} \frac{\lambda_i}{\lambda_0} \text{tr} K_i - \frac{1}{N-k} \sum_{i=0}^{n-1} \frac{\lambda_i}{\lambda_0} \text{tr} Y' K_i Y \right]^k \\ &\leq \left[\frac{1}{N-k} \alpha m_1 + \frac{1}{N-k} (n-m_1) - \frac{k}{N-k} \right]^k \\ &= \left(\alpha \left[\frac{m_1}{N-k} \right] + \left[\frac{N-m_1-k}{N-k} \right] \right)^k \end{aligned}$$

Die Ungleichungen sind vom gleichen Typ wie die für den Spezialfall $k = 1$, $n = 1$ gewonnenen Abschätzungen - wir können folglich zusammenfassend schließen, daß OLS-Schätzfunktionen β^* für β zwar Effizienzverluste gegenüber den GAUSS-MARKOV-Schätzern $\hat{\beta}$ aufweisen, daß schwerwiegender jedoch die möglicherweise erhebliche Unterschätzung der Varianz von β^* ist, die man mit der Verwendung der konventionellen Schätzfunktion für $\text{var}[\beta^*]$ riskiert.

3. SCHÄTZFUNKTIONEN FÜR DIE EIGENWERTE DER KOVARIANZMATRIX

3.0. VORBEMERKUNG

Das eingeführte Modell hat zwei verschiedene Parametersätze: die Regressionskoeffizienten β (Erwartungswertteil) und die Eigenwerte λ der Kovarianzmatrix (Kovarianzteil).

Die Abschätzungen des letzten Kapitels lassen erwarten, daß in der Regel die Verwendung der OLS-Schätzfunktionen β^* für β weniger problematisch sein wird als die Benutzung der unter Unabhängigkeitsannahmen gewonnenen "Schätzungen" für λ : $\hat{\lambda}_0 = \hat{\lambda}_1 = \dots = \hat{\lambda}_{n-1} = \hat{\sigma}^2$. Wir konzentrieren uns daher zunächst auf die Schätzung des Kovarianzteils. Es wird dabei durchgängig angenommen, daß $y \sim \mathcal{N}[XB, \lambda_0 K_0 + \lambda_1 K_1 + \dots + \lambda_{n-1} K_{n-1}]$.

3.1. ÜBERBLICK

Wir werden zwei alternative Strategien verwenden, um Schätzfunktionen für λ zu gewinnen. Das Schätzproblem und diese Strategien seien am Beispiel des Zweiebenenfalls illustriert.

Der Störfaktor u des Modells ist ein zufälliger Punkt des $\mathbb{R}^N$, von dem wir uns vorstellen können, daß er durch Überlagerung zweier orthogonaler Vektoren $u_0 := K_0 u$ und $u_1 := K_1 u$ (den "Mikrostörungen" und den "Makrostörungen") entstanden ist, also als Summe

$$u = u_0 + u_1$$

$$\text{mit } u_0 \in \mathcal{R}(K_0), u_1 \in \mathcal{R}(K_1) \text{ und } \mathcal{R}(K_0) \oplus \mathcal{R}(K_1) = \mathbb{R}^N.$$

Unser Ziel ist nun, mittels der Beobachtungen möglichst viel Informationen über die Verteilung dieser Komponenten zu gewinnen, d.h. also, da $u_1 \sim \mathcal{N}[0, \lambda_1 K_1]$ ist, den Parameter λ_1 der Kovarianzmatrix von u_1 zu schätzen. Dieses Ziel kann aufgrund zweier Tatsachen direkt nicht erreicht werden:

- (1) Es treten die Komponenten u_0 und u_1 nicht einzeln im Modell auf, sondern nur ihre Summe u .
- (2) Auch u ist nicht unmittelbar beobachtbar, sondern nur der Vektor der Beobachtungen y , der sich durch Überlagerung des Störvektors u mit einem festen, aber unbekannten Vektor $X\beta$ aus dem Spaltenraum $\mathcal{R}(X)$ von X ergibt.

Jeweils getrennt für sich sind diese beiden Hindernisse leicht zu überwinden:

- (1) wäre u ohne das Hinzutreten des Erwartungswertterms $X\beta$ beobachtbar, so könnten die Komponenten u_0 und u_1 durch Projektion in die jeweiligen Teilräume sofort aufgefunden werden. Ein Theorem von GRAYBILL und HULTQUIST [1961] zeigt, daß dann die quadratischen Formen

$$\hat{\lambda}_0 = \frac{u' K_0 u}{\text{tr } K_0} \quad \hat{\lambda}_1 = \frac{u' K_1 u}{\text{tr } K_1}$$

beste unverzerzte Schätzer für λ_0 und λ_1 liefern.

- (2) ginge es nur um eine Zerlegung von y in geschätzten "Erwartungswertterm" und Residuen, so liefert die bekannte Kleinstquadratetheorie "optimale" Projektionen P_y und Q_y mit $P := X(X'X)^{-1}X'$ und $Q := I - P$ in den Regressionsraum $\mathcal{R}(X)$ und dessen orthogonales Komplement $\mathcal{R}(X)^\perp$.

Das Schätzproblem wird dadurch interessant, daß beide Probleme: die Trennung der Mikro- und Makrobestandteile von u und die

Beseitigung des Erwartungswertteils von y gleichzeitig gelöst werden sollen, wobei jedoch die Operatoren K_0 (bzw. K_1) und P (bzw. Q), die eine "optimale" Lösung dieser Teilprobleme vermitteln, im allgemeinen nicht kommutieren - es also darauf ankommt, in welcher Reihenfolge man diese Operationen vornimmt.

Die erste der genannten Strategien geht von dem Residuenvektor $e = Qy$ aus und zerlegt diesen Vektor in Mikro- und Makroresiduen: $e_0 = K_0 Qy$ und $e_1 = K_1 Qy$. Die resultierenden Schätzer für λ_0 und λ_1 sind dann Durchschnitte dieser quadrierten Residuen ("ASR" - average of squared residuals).

Anstelle der OLS-Residuen Qy können mit alternativen a-priori-Annahmen über die Eigenwerte λ "GLS-Residuen" verwendet werden, die dann Generalisierte ASR (GASR) Schätzer liefern.

Sowohl ASR- als auch GASR-Schätzer sind im allgemeinen verzerrt - unter bestimmten Bedingungen existiert jedoch eine eindeutige lineare Transformation, die diese Schätzfunktionen in erwartungstreue überführt. Es läßt sich zeigen, daß die dann entstehenden Schätzfunktionen über die Erwartungstreue hinaus gewisse "lokale" Optimalitätseigenschaften aufweisen (MINQU, generalisierte MINQU-Schätzer).

Die zweite Strategie nimmt die Projektionen $y_0 = K_0 y$ und $y_1 = K_1 y$ in die Eigenräume von Σ zum Ausgangspunkt. Für diese Projektionen gelten die Partialmodelle

$$K_0 y = K_0 X \beta + K_0 u$$

und

$$K_1 y = K_1 X \beta + K_1 u.$$

Eine optimale Trennung in systematische und zufällige Teile der y_0 (bzw. y_1) kann innerhalb dieser Partialmodelle mit Hilfe der üblichen Kleinstquadratetheorie bewerkstelligt werden -

dabei werden jedoch Informationen zwischen den Partialmodellen unterdrückt.

Die resultierenden Schätzer für λ_0 und λ_1 (aus den Partialmodellen für $K_0 y$ bzw. für $K_1 y$) entsprechen bis auf eine kleine Korrektur den von BALESTRA (Anm. 1) vorgeschlagenen - seine Einführung dieser Schätzer als "global" beste unverzerrte quadratische Schätzfunktionen ist jedoch falsch.

3.2. ASR-SCHÄTZFUNKTIONEN

3.2.1. EINFÜHRUNG

In Abschnitt 1.3. wurde gezeigt, daß eine orthogonale Matrix

$$T = \begin{bmatrix} T_0 & | & T_1 & | & \dots & | & T_{n-1} \end{bmatrix}$$

$$(N,N) \quad (N,r_0) \quad (N,r_1) \quad \quad \quad (N,r_{n-1})$$

mit

$$T_s T_s' = K_s \quad (s = 0, 1, \dots, n-1)$$

und

$$r_s = \text{tr } K_s$$

existiert, so daß für alle Linearkombinationen

$$\Sigma = \lambda_0 K_0 + \dots + \lambda_{n-1} K_{n-1}$$

Anm. 1: BALESTRA [1973]. Das von BALESTRA betrachtete Modell ist formal unserem Zweiebenenmodell analog: BALESTRA betrachtet Kovarianzmatrizen, die die gewichtete Summe zweier bekannter idempotenter und gegenseitig orthogonaler Matrizen sind.

gilt, daß

$$T' \Sigma T = \begin{bmatrix} \lambda_0 I_0 & & & & \\ & \lambda_1 I_1 & & & \\ & & \ddots & & \\ & & & \ddots & \\ \emptyset & & & & \lambda_{n-1} I_{n-1} \end{bmatrix}$$

wobei die Matrix I_s die Dimension (r_s, r_s) besitzt. Im transformierten Modell

$$T'y = T'X\beta + T'u$$

hat $T'u$ folglich eine diagonale Varianz-Kovarianzmatrix der angegebenen Form - speziell gilt also, daß $T'_s u$ und $T'_t u$ unabhängig sind (für $s \neq t$; $s, t = 0 \dots n-1$) und daß

$$\text{var}[T'_s u] = \lambda_s I_s.$$

Bei beobachtbaren Störungen u würde man in diesem heteroskedastischen Modell λ_s durch

$$\frac{(u' T'_s)(T'_s u)}{r_s}$$

schätzen. Es liegt nun nahe, anstelle der in der Realität un beobachteten Störungen die OLS-Residuen

$$\begin{aligned} e &= y - X\beta^* \\ &= (I - X(X'X)^{-1}X')y = Qy \end{aligned}$$

zu benutzen und λ_s durch den Durchschnitt der r_s quadrierten Residuen $T'_s e$ (average of squared residuals)

$$\begin{aligned} \lambda_s^{(A)} &= \frac{(e' T'_s)(T'_s e)}{r_s} \\ &= \frac{(y - X\beta^*)' K_s (y - X\beta^*)}{r_s} \end{aligned}$$

zu schätzen.

3.2.2. FINITE EIGENSCHAFTEN DER ASR-SCHÄTZER

$\lambda_s^{(A)}$ ist eine quadratische Funktion der Elemente eines normalverteilten Zufallsvektors. Die folgenden Eigenschaften (Anm.1) derartiger quadratischer Formen werden im weiteren ständig benutzt.

SATZ 3.2.2.1. : Sei $y \sim W[\mu, \Sigma]$ mit Σ positiv definit vom Typ (N, N) - seien A, B symmetrische Matrizen vom Typ (N, N) , R eine beliebige (k, N) -Matrix.

Es gilt:

$$(1) \quad E[y' Ay] = \mu' A \mu + \text{tr } A \Sigma$$

$$(2) \quad \text{var}[y' Ay] = 2 \text{tr } A \Sigma A \Sigma + 4 \mu' A \Sigma A \mu$$

$$(3) \quad \text{cov}[y' Ay, y' By] = 2 \text{tr } A \Sigma B \Sigma + 4 \mu' A \Sigma B \mu$$

$$(4) \quad \text{cov}[Ry, y' Ay] = 2 R \Sigma A \mu$$

$$(5) \quad y' Ay \text{ und } y' By \text{ sind genau dann unabhängig, wenn } A \Sigma B = 0$$

$$(6) \quad y' Ay \text{ und } Ry \text{ sind genau dann unabhängig, wenn } R \Sigma A = 0$$

$$(7) \quad y' Ay \sim C[r, \frac{1}{2} \mu' A \mu] \text{ (Anm. 2) genau dann, wenn } A \Sigma \text{ idempotent vom Rang } r \text{ ist.}$$

Anm.1: siehe etwa SEARLE [1971], JOHNSON/KOTZ [1970]

Anm.2: Wir schreiben $s \sim C[r, \lambda]$, wenn s nichtzentral chiquadratverteilt ist mit r Freiheitsgraden und Nichtzentralitätsparameter λ , also die Verteilung einer Summe von r Quadraten unabhängiger

$W[\mu_j, 1]$ -verteilter Zufallsvariabler mit $\lambda = \frac{1}{2} \sum_{j=1}^r \mu_j^2$ besitzt.

Wir benutzen die Tatsache, daß die charakteristische Funktion $\phi(t)$ von s durch

$$\phi(t) = (1 - 2 i t) ^{-\frac{1}{2}} e^{\frac{i t \lambda}{1 - 2 i t}}$$

gegeben ist (i ist hier die imaginäre Einheit).

BEWEIS:

$$(1) \quad E[y'Ay] = \text{tr } AE[yy']$$

$$= \text{tr } A(\Sigma + \mu\mu')$$

$$= \mu'A\mu + \text{tr } A\Sigma$$

(2) Mit Σ ist auch Σ^{-1} positiv-definit. Es existiert daher eine nichtsinguläre (N,N) -Matrix P , so daß simultan gilt:

$$\Sigma^{-1} = PP'$$

$$A = PLP' \quad ,$$

wobei L eine Diagonalmatrix mit den (nicht notwendig verschiedenen) Eigenwerten $l_1, l_2, \dots, l_N$ von $\Sigma^{1/2} A \Sigma^{1/2}$ ist. (BELLMAN [1960, S. 58 ff]).

Es ist dann:

$$y'Ay = y'PLP'y = x'Lx \quad ,$$

wobei

$$x = P'y$$

gesetzt wurde. Man hat

$$E[x] = P'\mu =: \tau$$

und

$$\begin{aligned} \text{var}[x] &= P'\Sigma P = P'(PP')^{-1}P \\ &= P'(P')^{-1}P^{-1}P = I \quad , \end{aligned}$$

also

$$x \sim \mathcal{N}[\tau, I]$$

und

$$x_j^2 \sim C[1, \frac{1}{2}\tau_j^2]$$

für $j = 1, \dots, N$.

Die charakteristische Funktion von x_j^2 ist gleich

$$\phi_j(t) = (1-2it)^{-1/2} e^{\frac{it \tau_j^2}{1-2it}}$$

also die charakteristische Funktion von

$$x'x = \sum_{j=1}^N l_j x_j^2$$

gleich

$$\begin{aligned} \phi(t) &= \prod_{j=1}^N \phi_j(l_j t) \\ &= \prod_{j=1}^N (1-2it l_j)^{-1/2} e^{\frac{it \tau_j^2 l_j}{1-2it l_j}} \\ &= \left(\prod_{j=1}^N (1-2it l_j)^{-1/2} \right) e^{\sum_{j=1}^N \frac{it \tau_j^2 l_j}{1-2it l_j}} \\ &= \det(I-2it L)^{-1/2} e^{\tau'(it L)(I-2it L)^{-1} \tau} \end{aligned}$$

Schreiben wir θ für it , so ist also die kumulanten-erzeugende Funktion $\kappa(\theta)$ von $y'Ay = x'Lx$ gleich

$$\kappa(\theta) = \log \phi(t) = -1/2 \log \det(I-2\theta L) + \tau'(\theta L)(I-2\theta L)^{-1} \tau$$

Für θ hinreichend nahe bei 0 lassen sich $\log \det(I-2\theta L)$ und $L(I-2\theta L)^{-1}$ in elementweise konvergente Reihen um $\theta = 0$ entwickeln und man hat die übliche logarithmische Reihe

$$-1/2 \log \det(I-2\theta L) = \sum_{k=1}^{\infty} \frac{\theta^k}{k} 2^{k-1} (\text{tr } L^k)$$

und die Neumann-Reihe

$$\Theta L(I - 2\Theta L)^{-1} = \sum_{k=1}^{\infty} 2^{k-1} \Theta^k L^k$$

woraus sich

$$\begin{aligned} \kappa(\Theta) &= \sum_{k=1}^{\infty} \frac{\Theta^k}{k} 2^{k-1} [(\text{tr } L^k) + k \tau' L^k \tau] \\ &= \sum_{k=1}^{\infty} \frac{\Theta^k}{k!} (k-1)! 2^{k-1} [(\text{tr } L^k) + k \tau' L^k \tau] \end{aligned}$$

ergibt.

Bezeichnen wir die k 'te Kumulante von $y'Ay$ mit $\kappa_k(y'Ay)$ und beachten

$$\text{tr } L^k = \text{tr } (\Sigma^{1/2} A \Sigma^{1/2})^k = \text{tr } (A \Sigma)^k$$

sowie

$$\tau' L^k \tau = \mu' P L^k P' \mu = \mu' A (\Sigma A)^{k-1} \mu$$

so gilt

$$\kappa_k(y'Ay) = (k-1)! 2^{k-1} (\text{tr } (A \Sigma)^k) + k \mu' A (\Sigma A)^{k-1} \mu$$

und speziell für die Varianz

$$\kappa_2(y'Ay) = \text{var}(y'Ay) = 2 \text{tr } A \Sigma A \Sigma + 4 \mu' A \Sigma A \mu$$

(3) Es ist (Polarisierung!)

$$\text{cov}[y'Ay, y'By] = 1/2(\text{var}[y'(A+B)y] - \text{var}[y'Ay] - \text{var}[y'By])$$

Mit

$$\begin{aligned} \text{var}[y'(A+B)y] &= 2 \text{tr}(A+B)\Sigma(A+B)\Sigma + 4 \mu'(A+B)\Sigma(A+B)\mu \\ &= 2 \text{tr } A \Sigma A \Sigma + 2 \text{tr } B \Sigma B \Sigma + 4 \text{tr } A \Sigma B \Sigma + 4 \mu' A \Sigma A \mu \\ &\quad + 4 \mu' B \Sigma B \mu + 8 \mu' A \Sigma B \mu \\ &= \text{var}[y'Ay] + \text{var}[y'By] + 4 \text{tr } A \Sigma B \Sigma + 8 \mu' A \Sigma B \mu \end{aligned}$$

folgt die Behauptung.

(4) Betrachten wir zunächst

$$\begin{aligned}
 \text{cov}[y, y' Ay] &= E[(y - \mu)(y' Ay - E[y' Ay])] \\
 &= E[(y - \mu)(y' Ay - \mu' A \mu - \text{tr } A \Sigma)] \\
 &= E[(y - \mu)((y - \mu)' A (y - \mu) + 2 y' A \mu \\
 &\quad - 2 \mu' A \mu - \text{tr } A \Sigma)] \\
 &= E[(y - \mu)(y - \mu)' A (y - \mu)] + 2 E[(y - \mu)(y - \mu)' A \mu] \\
 &\quad - (\text{tr } EA) E[y - \mu] .
 \end{aligned}$$

Da $y_j - \mu_j$ ($j = 1 \dots N$) normalverteilt ist mit Erwartungswert 0, gilt für die dritten Momente

$$E[(y_i - \mu_i)(y_j - \mu_j)(y_k - \mu_k)] = 0 \quad (i, j, k = 1 \dots N)$$

und man hat

$$\text{cov}[y, y' Ay] = 2 E[(y - \mu)(y - \mu)' A \mu] = 2 \Sigma A \mu .$$

Es folgt

$$\text{cov}[Ry, y' Ay] = R \text{cov}[y, y' Ay] = 2 R \Sigma A \mu .$$

(5) Seien $y' Ay$ und $y' By$ unabhängig, dann gilt

$$\text{cov}[y' Ay, y' By] = 2 \text{tr } A \Sigma B \Sigma + 4 \mu' A \Sigma B \mu = 0$$

woraus $A \Sigma B = 0$ folgt (die Beziehung muß für alle μ gelten!).

Andererseits sei $A \Sigma B = 0$, dann ist

$$\text{cov}[Ay, By] = A \Sigma B = 0 ,$$

folglich Ay und By unabhängig. Seien A^- , B^- generalisierte Inverse von A bzw. B (also Matrizen A^- mit $AA^-A = A$ und B^- mit $BB^-B = B$), dann sind

$$y' Ay = (y' A) A^- (Ay)$$

und

$$y' By = (y' B) B^- (By)$$

Funktionen der unabhängigen Zufallsvariablen A_y und B_y , also selbst unabhängig. (Anm. 1)

(6) Beweis ebenso wie Teil (5).

(7) Sei $A \in$ idempotent vom Rang r , dann sind die Eigenwerte von $\Sigma^{1/2} A \Sigma^{1/2}$ 1 bzw. 0 mit den Vielfachheiten r bzw. $(N-r)$ und L läßt sich in der Form

$$L = \begin{bmatrix} I & 0 \\ 0 & 0 \end{bmatrix}$$

mit I vom Typ (r, r) schreiben. Die charakteristische Funktion von $y'Ay$ ist dann

$$\begin{aligned} \phi(t) &= \left(\prod_{j=1}^r (1-2it)^{-1/2} \right) e^{j \sum_{j=1}^r \frac{it \tau_j^2}{1-2it}} \\ &= (1-2it)^{-r/2} e^{\frac{it}{1-2it} \sum_{j=1}^r \tau_j^2} \end{aligned}$$

Dies ist die charakteristische Funktion einer $C[r, \lambda]$ -verteilten Zufallsvariable

$$\lambda = 1/2 \sum_{j=1}^r \tau_j^2 = 1/2 \tau' L \tau = 1/2 \mu' A \mu.$$

Sei andererseits $y'Ay \sim C[r, 1/2 \mu' A \mu]$, dann lautet die charakteristische Funktion von $y'Ay$

Anm. 1: Der Beweis von SEARLE [1971, S.60] ist ohne die Einführung komplexer multivariater Normalverteilungen unrichtig. (A und B sind nicht notwendig positiv-semidefinit!)

$$\phi(t) = (1-2it)^{-r/2} e^{\frac{it}{1-2it}} {}^2 (1/2 \mu' A \mu)$$

Nehmen wir o.B.d.A. an, daß $\mu = 0$, so wird

$$\phi(t) = (1-2it)^{-r/2}$$

Wir suchen Bedingungen für die Eigenwerte l_j von $\Sigma^{1/2} A \Sigma^{1/2}$, so daß

$$(1-2it)^{-r/2} = \prod_{j=1}^N (1-2it l_j)^{-1/2}$$

Es müssen offenbar r der l_j gleich 1, die anderen $(N-r)$ Eigenwerte gleich 0 sein, woraus

$$L = \begin{bmatrix} I & 0 \\ 0 & 0 \end{bmatrix}$$

und daraus weiter Idempotenz von $\Sigma^{1/2} A \Sigma^{1/2}$ und $A \Sigma$ folgt.

Um die Aussagen des folgenden Satzes kompakt darstellen zu können, definieren wir

$$\lambda(A) := (\lambda_0^{(A)}, \lambda_1^{(A)}, \dots, \lambda_{n-1}^{(A)}),$$

$$Q := I - X(X'X)^{-1}X'$$

$$V := (v_{ij}) = \left(\frac{\text{tr} K_i Q K_j Q}{\text{tr} K_i} \right) \quad (i, j = 0 \dots n-1)$$

$$v^{(s,t)} := (v_{ij}^{(s,t)}) = (\text{tr} K_s Q K_i Q K_t Q K_j Q)$$

SATZ 3.2.2.2. : Es gilt

(1) $\lambda_s^{(A)}$ ist eine nichtnegative quadratische Form in y :

$$\lambda_s^{(A)} = y' \begin{bmatrix} Q K_s Q \\ r_s \end{bmatrix} y$$

- (2) $\lambda_s^{(A)}$ ist translationsinvariant, d.h.

$$\lambda_s^{(A)} = (y - Xb)' \begin{bmatrix} QK_s Q \\ r_s \end{bmatrix} (y - Xb)$$

für alle $b \in \mathbb{R}^k$.

- (3) $\lambda^{(A)}$ ist im allgemeinen verzerrt:

$$E[\lambda^{(A)}] = V\lambda$$

$$(4) \quad \text{var}[\lambda_s^{(A)}] = \frac{2\lambda' V^{(s,s)} \lambda}{r_s^2}$$

$$(5) \quad \text{cov}[\lambda_s^{(A)}, \lambda_t^{(A)}] = \frac{2\lambda' V^{(s,t)} \lambda}{r_s r_t}$$

- (6) $\lambda^{(A)}$ und β^* sind im allgemeinen nicht unabhängig, es gilt jedoch

$$\text{cov}[\lambda^{(A)}, \beta^*] = 0.$$

BEWEIS:

- (1) Man hat

$$\lambda_s^{(A)} = (y - Xb^*)' \begin{bmatrix} K_s \\ r_s \end{bmatrix} (y - Xb^*) = y' \begin{bmatrix} QK_s Q \\ r_s \end{bmatrix} y$$

und $QK_s Q$ ist wegen $K_s = T_s T_s'$ positiv semidefinit.

- (2) Klar wegen $QX = 0$.

- (3) Aus Satz 3.2.2.1. (1) folgt

$$\begin{aligned} E[\lambda_s^{(A)}] &= \text{tr} \left[\frac{QK_s Q}{r_s} \right] + \beta' X' \begin{bmatrix} QK_s Q \\ r_s \end{bmatrix} X\beta \\ &= \sum_{t=C}^{n-1} \left(\frac{\text{tr } QK_s Q K_t}{\text{tr } K_s} \right) \lambda_t \end{aligned}$$

also

$$E[\lambda^{(A)}] = V\lambda.$$

(4) Aus Satz 3.2.2.1. (2) folgt

$$\begin{aligned} \text{var}[\lambda_s^{(A)}] &= 2 \text{tr} \left[\frac{QK_s Q}{r_s} \Sigma \frac{QK_s Q}{r_s} \Sigma \right] \\ &= 2 \sum_{i=0}^{n-1} \sum_{j=0}^{n-1} \text{tr} \frac{QK_s QK_i QK_s QK_j}{r_s^2} \lambda_i \lambda_j \\ &= \frac{2 \lambda' V(s, s) \lambda}{r_s^2} \end{aligned}$$

(5) Aus Satz 3.2.2.1. (3)

(6) Es ist

$$\begin{aligned} \text{cov}[\beta_s^*, \lambda_s^{(A)}] &= \text{cov}[(X'X)^{-1} X' y, y' \frac{QK_s Q}{r_s} y] \\ &= \frac{2}{r_s} (X'X)^{-1} X' \Sigma QK_s Q X \beta \\ &= 0 \end{aligned}$$

wegen

$$QX = (I - X(X'X)^{-1} X')X = 0;$$

jedoch im allgemeinen

$$(X'X)^{-1} X' \Sigma QK_s Q \neq 0.$$

Der Ausdruck für den Erwartungswert von $\lambda_s^{(A)}$

$$E[\lambda_s^{(A)}] = \sum_{t=0}^{n-1} \frac{\text{tr} QK_s QK_t}{\text{tr} K_s} \lambda_t$$

kann instruktiver

$$E[\lambda_s^{(A)}] = (1 - 2 \frac{\text{tr}(X'X)^{-1} X' K_s X}{\text{tr} K_s}) \lambda_s + \sum_{t=0}^{n-1} (\frac{\text{tr}(X'X)^{-1} X' K_s X (X'X)^{-1} X' K_t X}{\text{tr} K_s}) \lambda_t$$

geschrieben werden - man erkennt so, daß der Bias von $\lambda_s^{(A)}$ von zwei Faktoren abhängt:

- (1) die Existenz von "Mehrebenenwirkungen" in X , also der Größe des Terms $\text{tr}(X'X)^{-1} X' K_s X$
- (2) die Ähnlichkeit der Mehrebenenzerlegungen $X' K_s X$ und $X' K_t X$, die durch das innere Produkt $\text{tr}(X'X)^{-1} X' K_s X (X'X)^{-1} X' K_t X$ angezeigt wird.

Der Einfluß beider Faktoren wird kleiner, wenn $\text{tr} K_s$, also die Differenz der Aggregate in den Zerlegungen $\mathcal{Z}_s$ und $\mathcal{Z}_{s+1}$, erhöht wird.

3.2.3. ASYMPTOTISCHE EIGENSCHAFTEN

Bei der asymptotischen Analyse von Regressionsmodellen mit abhängigen Beobachtungen entsteht das Problem, daß bei einer Hinzufügung von Beobachtungen zu einer gegebenen Stichprobe die Kovarianzmatrix des Samples auf mehr als eine Weise verändert werden kann.

So kann etwa in einer Zweiebenenstruktur durch Vermehrung der Beobachtungen

- (1) die Zahl der Beobachtungen innerhalb der einzelnen Gruppen bei gleichbleibender Gruppenzahl

oder

(2) die Anzahl der Gruppen bei gleichbleibender Gruppengröße

erhöht werden. Eine Struktur mit mehr als zwei Ebenen erlaubt eine noch größere Vielfalt von Weisen, den Baum, durch den die Mehrebenenstruktur definiert wird, "wachsen" zu lassen.

Die Betrachtung der Ausdrücke für Erwartungswert und Varianz der ASR-Schätzer legt es nun nahe, nur solche Stichproben als 'groß' zu betrachten, in denen alle Terme $r_s = \text{tr } K_s$ für $s = 0, 1, \dots, n-1$ 'groß' sind - in denen also, anschaulich gesprochen, die Zahl der "Zweige" auf allen Ebenen des Baumes groß ist. Wir betrachten dementsprechend Folgen von Regressionsproblemen, bei denen die Terme r_s über alle Grenzen wachsen, wobei dieses Wachstum sich jedoch in "gleichmäßiger Weise" vollziehen soll, d.h. r_s und r_t für $s, t = 0, \dots, n-1$ asymptotisch äquivalente Folgen sein sollen.

Gegeben sei also eine Folge von Regressionsproblemen, also eine Folge von $(n+2)$ -Tupeln

$$\begin{aligned} & [y, X, K_0, \dots, K_{n-1}]_{N_0} \\ & [y, X, K_0, \dots, K_{n-1}]_{N_0+1} \\ & \vdots \\ & [y, X, K_0, \dots, K_{n-1}]_N \\ & \vdots \end{aligned}$$

in der für die $(n+2)$ Elemente $y_{(N)}, X_{(N)}, K_{0(N)}, \dots, K_{n-1(N)}$ des N 'ten Tupels ($N = N_0, N_0+1, \dots$) folgende Annahmen gelten:

- (1) $X_{(N)}$ ist eine nichtstochastische (N, k) -Matrix mit $\text{rg}(X_{(N)}) = k < N_0 \leq N$

- (2) $K_{s(N)}$ ($s=0,1,\dots,n-1$) ist eine nichtstochastische symmetrische (N,N) -Matrix mit

$$\begin{aligned} K_{s(N)} K_{s(N)} &= K_{s(N)} \\ K_{s(N)} K_{t(N)} &= 0 \quad (s \neq t) \\ K_{0(N)} + \dots + K_{n-1(N)} &= I \\ \operatorname{rg}(K_{s(N)}) = \operatorname{tr}(K_{s(N)}) = r_{s(N)} &> 0 \quad . \end{aligned}$$

- (3) $y_{(N)}$ ist ein multivariat normalverteilter Zufallsvektor mit

$$\begin{aligned} E[y_{(N)}] &= x_{(N)} \beta \\ \operatorname{var}[y_{(N)}] &= \lambda_0 K_{0(N)} + \lambda_1 K_{1(N)} + \dots + \lambda_{n-1} K_{n-1(N)} \quad . \end{aligned}$$

Dabei sind β und $\lambda = (\lambda_0, \lambda_1, \dots, \lambda_{n-1})'$ nichtstochastische und nicht von N abhängende k bzw. n -dimensionale vektorwertige Parameter mit

$$\begin{aligned} \beta \in \Omega_1 &= R^k \\ \lambda \in \Omega_2 &= R_+^n \quad (\text{Anm. 1}) \quad . \end{aligned}$$

Der Parameterraum des Modells ist also das kartesische Produkt $\Omega = \Omega_1 \times \Omega_2$.

Über das asymptotische Verhalten dieser Folge werden folgende Annahmen getroffen:

- (4) Für alle $s=0,1,\dots,n-1$ divergieren die Folgen $\{r_{s(N)}\}$ bestimmt:

$$\lim_{N \rightarrow \infty} r_{s(N)} = \infty$$

wobei

$$\lim_{N \rightarrow \infty} \frac{r_{s(N)}}{N} = k_s \in (0,1) \quad .$$

Anm. 1: $R_+^n = \{z = (z_1, z_2, \dots, z_n)' \mid z_i > 0 \text{ für } i=1,\dots,n\}$

Die Folgen $(r_0)_N, \dots, (r_{n-1})_N, (N)$ sind also sämtlich asymptotisch äquivalent.

(5) Für alle $s=0, 1, \dots, n-1$ existieren die Grenzwertmatrizen

$$\lim_{N \rightarrow \infty} \left[\frac{1}{r_s(N)} X'_{(N)} K_s(N) X_{(N)} \right] = S_s$$

und seien endlich.

(6) Sei $\lambda = (\lambda_0, \lambda_1, \dots, \lambda_{n-1})' \in \Omega_2$, dann existiert für

$$\Sigma_{(N)} := \lambda_0 K_0(N) + \dots + \lambda_{n-1} K_{n-1}(N)$$

sowohl

$$\lim_{N \rightarrow \infty} \left[\frac{1}{N} X'_{(N)} \Sigma_{(N)}^{-1} X_{(N)} \right] = \sum_{s=0}^{n-1} \frac{\lambda_s}{\lambda_s} S_s$$

als auch

$$\lim_{N \rightarrow \infty} \left[\frac{1}{N} X'_{(N)} \Sigma_{(N)}^{-1} X_{(N)} \right]^{-1}.$$

Beide Grenzmatriizen seien endlich und es gelte

$$\left[\lim_{N \rightarrow \infty} \left[\frac{1}{N} X'_{(N)} \Sigma_{(N)}^{-1} X_{(N)} \right]^{-1} \right]^{-1} = \lim_{N \rightarrow \infty} \left[\frac{1}{N} X'_{(N)} \Sigma_{(N)}^{-1} X_{(N)} \right]$$

Annahmen (5) und (6) besagen wenig mehr als die gleichmäßige Beschränktheit von $X_{(N)}$ - es werden also Folgen von Regressormatrizen ausgeschlossen, deren Elemente über alle Grenzen wachsen.

Wir befreien uns im folgenden wieder von den lästigen Indizes (N) - Aussagen, die der Bequemlichkeit halber für "alle" oder "beliebige" $(y, X, K_0, \dots, K_{n-1})$ ausgesprochen werden, sind so zu verstehen, daß deren Gültigkeit für "fast alle" Glieder der Folge behauptet wird.

Grundlegend für die Untersuchung der asymptotischen Eigenschaften der ASR-Schätzer $\lambda_s^{(A)}$ ist die folgende Zerlegung der Einheits-

matrix.

Man hat

$$X'X = X'K_0X + X'K_1X + \dots + X'K_{n-1}X \quad .$$

Da $X'X$ positiv-definit ist, existiert eine eindeutige symmetrische, positiv-definite Quadratwurzel von $(X'X)^{-1} : (X'X)^{-1/2}$ und es gilt

$$I_k = (X'X)^{-1/2} X'K_0 X (X'X)^{-1/2} + \dots + (X'X)^{-1/2} X'K_{n-1} X (X'X)^{-1/2}$$

Schreiben wir C_s ($s=0,1,\dots,n-1$) für die Matrix $(X'X)^{-1/2} X'K_s X (X'X)^{-1/2}$, so gilt also

$$I = C_0 + C_1 + \dots + C_{n-1} \quad .$$

SATZ 3.2.3.1. : Man hat für $s,t=0,1,\dots,n-1$

- (1) $0 \leq \text{tr } C_s \leq k$
- (2) $0 \leq \text{tr } C_s^2 \leq \text{tr } C_s$
- (3) Die Matrix $F := (f_{st}) := (\text{tr } C_s C_t)$ ist positiv-semidefinit.
- (4) $0 \leq \text{tr } C_s C_t \leq \min (\text{tr } C_s, \text{tr } C_t)$

BEWEIS :

- (1) Da C_s positiv definit ist, gilt

$$\text{tr } C_s \geq 0 \quad .$$

Die andere Seite der Ungleichung folgt dann aus

$$\sum_{s=0}^{n-1} \text{tr } C_s = \text{tr } I_k = k \quad .$$

(2) Es ist

$$C_s = Z'K_sZ$$

wobei

$$Z := X(X'X)^{-1/2}$$

die Bedingung $Z'Z = I$ erfüllt. Der POINCARÉ'sche Einschließungssatz (siehe etwa RAO [1973, S.64]) zeigt dann für den größten Eigenwert $\lambda_{\max}(C_s)$ von C_s , daß

$$\lambda_{\max}(C_s) \leq \lambda_{\max}(K_s) = 1$$

und für den kleinsten Eigenwert $\lambda_{\min}(C_s)$ von C_s , daß

$$\lambda_{\min}(C_s) \geq \lambda_{\min}(K_s) = 0.$$

Alle Eigenwerte von C_s liegen folglich zwischen 0 und 1, woraus

$$\text{tr } C_s^2 \leq \text{tr } C_s$$

folgt.

Die andere Seite der Ungleichung ist klar.

(3) Schreiben wir

$$C = [\text{vec } C_0 | \text{vec } C_1 | \dots | \text{vec } C_{n-1}] ,$$

so ist

$$F = (f_{st}) = (\text{tr } C_s C_t) = C' C$$

positiv semidefinit.

(4) $\text{tr } C_s C_t \geq 0$ gilt wegen positiver Semidefinitheit von C_s und C_t .

Die andere Seite der Ungleichung folgt dann aus

$$\sum_{s=0}^{n-1} \text{tr } C_s C_t = \text{tr } C_t$$

und

$$\sum_{t=0}^{n-1} \text{tr } C_s C_t = \text{tr } C_s .$$

Wir benutzen die Beschränktheit der C_s um im folgenden einen Satz über asymptotische Eigenschaften der Kumulanten von $\lambda_s^{(A)}$ abzuleiten:

SATZ 3.2.3.2. : Sei $\kappa_h(\lambda_s^{(A)})$ die h 'te Kumulante von $\lambda_s^{(A)}$. Dann gilt

$$\kappa_h(\lambda_s^{(A)}) = (h-1)! 2^{h-1} \frac{\lambda_s^h}{r_s^{h-1}} + O\left(\frac{1}{r_s^h}\right) .$$

Speziell also

$$\begin{aligned} E[\lambda_s^{(A)}] &= \kappa_1(\lambda_s^{(A)}) = \lambda_s + O\left(\frac{1}{r_s}\right) \\ \text{var } \lambda_s^{(A)} &= \kappa_2(\lambda_s^{(A)}) = 2 \frac{\lambda_s^2}{r_s} + O\left(\frac{1}{r_s^2}\right) . \end{aligned}$$

BEWEIS : Im Beweis von Satz 3.2.2.1. (2) wurde allgemein gezeigt, daß für $y \sim \mathcal{W}[\mu, \Sigma]$ die h 'te Kumulante von $y'Ay$ durch

$$\kappa_h(y'Ay) = (h-1)! 2^{h-1} [\text{tr}(A\Sigma)^h + h \mu'A(\Sigma A)^{h-1} \mu]$$

gegeben ist. Hier ist

$$\Sigma = \lambda_C K_C + \dots + \lambda_{n-1} K_{n-1}$$

$$A = r_s^{-1} \mathcal{Q} K_s \mathcal{Q} \quad \text{mit} \quad \mathcal{Q} = I - F = I - X(X'X)^{-1}X'$$

$$\mu = X\beta$$

so daß

$$\begin{aligned}\kappa_h(\lambda_s^{(A)}) &= (h-1)! 2^{h-1} \operatorname{tr} \left[\frac{QK_s Q\Sigma}{r_s} \right]^h \\ &= \frac{(h-1)! 2^{h-1}}{r_s^h} \operatorname{tr} [QK_s Q\Sigma]^h\end{aligned}$$

Nun ist

$$\operatorname{tr}[QK_s Q\Sigma]^h = \sum_{t_1=0}^{n-1} \sum_{t_2=0}^{n-1} \dots \sum_{t_h=0}^{n-1} \left\{ \operatorname{tr} \left[\prod_{p=1}^h QK_s QK_{t_p} \right] \prod_{p=1}^h \lambda_{t_p} \right\} .$$

Wir entwickeln den Ausdruck

$$\operatorname{tr} \left[\prod_{p=1}^h (I-P) K_s (I-P) K_{t_p} \right]$$

in P und finden die Summe von 4^h Termen der Form

$$\operatorname{tr} [\bar{P} K_s \bar{P} K_{t_1} \bar{P} K_s \bar{P} K_{t_2} \dots \bar{P} K_s \bar{P} K_{t_h}]$$

wobei

$$\bar{P} \in \{I, -P\} .$$

Der Term

$$\operatorname{tr} [K_s K_{t_1} K_s K_{t_2} \dots K_s K_{t_n}]$$

ist nur dann ungleich 0, wenn

$$t_1 = t_2 = \dots = t_h = s .$$

Alle anderen Terme sind von der Form

$$\operatorname{tr} [C_{\alpha_1} C_{\alpha_2} \dots C_{\alpha_g}]$$

mit $g \in \{1, 2, \dots, 2h\}$ und α_j ($j=1 \dots g$) $\in \{s, t_1, \dots, t_h\}$ und bleiben folglich aufgrund der CAUCHY-SCHWARZ'schen Ungleichung und Satz 3.2.3.1. beschränkt, wenn r_s bestimmt divergiert.

Es ergibt sich somit für alle h und $r_s \rightarrow \infty$

$$\text{tr}[\mathbf{Q}\mathbf{K}_s\mathbf{Q}\Sigma]^h = (\text{tr}[\mathbf{K}_s])\lambda_s^h + o(1)$$

und damit

$$\kappa_h(\lambda_s^{(A)}) = (h-1)! 2^{h-1} \frac{\lambda_s^h}{r_s^{h-1}} + o\left(\frac{1}{r_s^h}\right).$$

Mit den Ergebnissen dieses Satzes können wir asymptotische Eigenschaften der ASR-Schätzfunktionen gewinnen:

SATZ 3.2.3.3. : Es gilt:

- (1) $\lambda_s^{(A)}$ ist asymptotisch unverzerrt
- (2) $\lambda_s^{(A)}$ ist eine konsistente Schätzfunktion für λ_s
- (3) $\sqrt{r_s} \lambda_s^{(A)}$ und $\sqrt{r_t} \lambda_t^{(A)}$ sind asymptotisch unabhängig für $s \neq t$
- (4) $\sqrt{r_s}(\lambda_s^{(A)} - \lambda_s)$ ist asymptotisch normalverteilt mit Erwartungswert 0 und Varianz $2\lambda_s^2$

BEWEIS :

- (1) Man hat mit dem letzten Satz

$$\text{Bias}(\lambda_s^{(A)}) = (E[\lambda_s^{(A)}] - \lambda_s)^2 = o\left(\frac{1}{r_s^2}\right)$$

also

$$\lim_{r_s \rightarrow \infty} (E[\lambda_s^{(A)}] - \lambda_s)^2 = 0.$$

- (2) Wir zeigen, daß

$$\lim_{r_s \rightarrow \infty} \lambda_s^{(A)} = \lambda_s,$$

also

$$\lim_{r_s \rightarrow \infty} E[\lambda_s^{(A)} - \lambda_s]^2 = 0.$$

Nun gilt die Zerlegung

$$E[\lambda_s^{(A)} - \lambda_s]^2 = (E[\lambda_s^{(A)}] - \lambda_s)^2 + \text{var}[\lambda_s^{(A)}] .$$

In Teil (1) wurde bereits

$$\lim_{r_s \rightarrow \infty} (E[\lambda_s^{(A)}] - \lambda_s)^2 = 0$$

gezeigt, mit Satz 3.2.3.2. findet man

$$\text{var}[\lambda_s^{(A)}] = \frac{2\lambda_s^2}{r_s} + O\left(\frac{1}{r_s^2}\right)$$

also

$$\lim_{r_s \rightarrow \infty} \text{var}[\lambda_s^{(A)}] = 0$$

und damit

$$\lim_{r_s \rightarrow \infty} \lambda_s^{(A)} = \lambda_s .$$

- (3) Es gilt, daß $y'A_y$ und $y'By$ genau dann unabhängig sind, wenn $A\Sigma B = 0$.

Wir zeigen, daß diese Beziehung für

$$A = \frac{QK_s Q}{\sqrt{r_s}} \quad \text{und} \quad B = \frac{QK_t Q}{\sqrt{r_t}}$$

asymptotisch erfüllt ist.

Es ist $A\Sigma B = 0$ genau dann, wenn

$$\text{tr}(A\Sigma B^2 \Sigma A) = 0 .$$

Nun findet man

$$\text{tr}(A\Sigma B^2 \Sigma A) = \frac{1}{r_s r_t} \sum_{i=0}^{n-1} \sum_{j=0}^{n-1} [\text{tr}(I-P)K_s(I-P)K_i(I-P)K_t(I-P)K_t(I-P)K_j(I-P)K_t(I-P)K_t] \lambda_i \lambda_j$$

mit $P = X(X'X)^{-1}X'$,

und man stellt fest, daß wegen $K_s K_t = 0$, ebenso

wie im Beweis des letzten Satzes alle Terme in der Summe beschränkt sind - es existiert kein Term der Form $\text{tr}(K_s)$.

Also

$$\text{tr}(A \Sigma B^2 \Sigma A) = \frac{1}{r_s r_t} O(1)$$

weshalb

$$\lim_{\substack{r_s \rightarrow \infty \\ r_t \rightarrow \infty}} \text{tr}(A \Sigma B^2 \Sigma A) = 0$$

und

$$\lim_{\substack{r_s \rightarrow \infty \\ r_t \rightarrow \infty}} A \Sigma B = 0 .$$

(4) Der Erwartungswert von $\sqrt{r_s}(\lambda_s^{(A)} - \lambda_s)$ ist $O(\frac{1}{\sqrt{r_s}})$, die Varianz ist

$$\text{var}[\sqrt{r_s}(\lambda_s^{(A)} - \lambda_s)] = 2\lambda_s^2 + O(\frac{1}{r_s}) .$$

Für die h'te Kumulante findet man

$$\kappa_h[\sqrt{r_s}(\lambda_s^{(A)} - \lambda_s)] = \frac{\lambda_s^h (h-1)! 2^{h-1}}{r_s^{(\frac{h}{2}-1)}} + O(\frac{1}{r_s^{\frac{h}{2}}}) .$$

Es gilt also

$$\lim_{r_s \rightarrow \infty} E[\sqrt{r_s}(\lambda_s^{(A)} - \lambda_s)] = 0 ,$$

$$\lim_{r_s \rightarrow \infty} \text{var}[\sqrt{r_s}(\lambda_s^{(A)} - \lambda_s)] = 2\lambda_s^2 ,$$

$$\lim_{r_s \rightarrow \infty} \kappa_h[\sqrt{r_s}(\lambda_s^{(A)} - \lambda_s)] = 0 \quad (h = 3, 4, \dots)$$

Die Folge dieser Grenzkumulanten definiert eindeutig eine Grenzverteilung - die Normalverteilung mit Erwartungswert 0 und Varianz $2\lambda_s^2$.

Approximative Aussagen über die Verteilung der $\lambda_s^{(A)}$ können also für "große" r_s aus der Tatsache gewonnen werden, daß

$$\begin{bmatrix} \lambda_0^{(A)} \\ \lambda_1^{(A)} \\ \vdots \\ \lambda_{n-1}^{(A)} \end{bmatrix} \sim \text{approx. } \mathcal{N} \left(\begin{bmatrix} \lambda_0 \\ \lambda_1 \\ \vdots \\ \lambda_{n-1} \end{bmatrix}, \begin{bmatrix} \frac{2\lambda_0^2}{r_0} & & & \\ & \frac{2\lambda_1^2}{r_1} & & \\ & & \ddots & \\ & & & \frac{2\lambda_{n-1}^2}{r_{n-1}} \end{bmatrix} \right)$$

Bei der Untersuchung der Likelihoodfunktion des Modells (Kapitel 4) werden wir erkennen, daß es sich bei den $\lambda_s^{(A)}$ nicht nur um konsistente und asymptotisch normalverteilte Schätzer handelt, sondern sogar um BAN-Schätzfunktionen, also Beste Asymptotisch Normalverteilte Schätzer.

3.3. GENERALISIERTE ASR-SCHÄTZER

3.3.1. EINFÜHRUNG

Wir betrachten wieder Schätzfunktionen des Typs

$$\lambda_s^{(GA)} = \frac{(y - X\beta^*)' K_s (y - X\beta^*)}{r_s}$$

wobei hier jedoch für β^* nicht die OLS-Schätzfunktion verwendet wird, sondern eine "GLS"-Schätzfunktion

$$\beta^* = (X' \Sigma_0^{-1} X)^{-1} X' \Sigma_0^{-1} y$$

mit

$$\Sigma_0 = \lambda_0^{(0)} K_0 + \lambda_1^{(0)} K_1 + \dots + \lambda_{n-1}^{(0)} K_{n-1},$$

wobei die $\lambda_t^{(0)}$ ($t = 0, 1, \dots, n-1$) vorgegebene Konstante aus $\mathbb{R}^+$ sind.

Die Untersuchung dieser Schätzfunktion ist notwendig, um Situationen zu erfassen, in denen Vorkenntnisse über die λ_s (genauer: über deren Verhältnisse) vorliegen. Unsere Vorgehensweise wird dabei ganz analog zur Untersuchung der speziellen ASR-Schätzfunktion mit $\Sigma_0 = I$ sein, so daß die Darstellung sehr kurz sein kann.

3.3.2. FINITE EIGENSCHAFTEN

Wir definieren

$$Q_0 := I - X(X'\Sigma_0^{-1}X)^{-1}X'\Sigma_0^{-1}$$

$$V_0 := (v_{st}) = \left(\frac{\text{tr } Q_0' K_s Q_0 K_t}{\text{tr } K_s} \right)$$

$$v_0^{(s,t)} := (v_{ij}^{(s,t)}) = (\text{tr } Q_0' K_s Q_0 K_i Q_0' K_t Q_0 K_j)$$

$$\lambda^{(GA)} := (\lambda_0^{(GA)}, \lambda_1^{(GA)}, \dots, \lambda_{n-1}^{(GA)})$$

SATZ 3.3.2.1. : Es gilt

(1) $\lambda_s^{(GA)}$ ist eine nichtnegative quadratische Form in y

$$\lambda_s^{(GA)} = y' \left[\frac{Q_0' K_s Q_0}{r_s} \right] y$$

- (2) $\lambda_s^{(GA)}$ ist translationsinvariant, d.h. für alle $b \in \mathbb{R}^k$ ist

$$\lambda_s^{(GA)} = (y - Xb)' \left[\frac{Q' K_s Q}{r_s} \right] (y - Xb)$$

- (3) $\lambda_s^{(GA)}$ ist im allgemeinen verzerrt:

$$E[\lambda_s^{(GA)}] = v_o \lambda$$

$$(4) \quad \text{var}[\lambda_s^{(GA)}] = \frac{2\lambda' v_o^{(s,s)} \lambda}{r_s^2}$$

$$(5) \quad \text{cov}[\lambda_s^{(GA)}, \lambda_t^{(GA)}] = \frac{2\lambda' v_o^{(s,t)} \lambda}{r_s r_t}$$

- (6) $\lambda^{(GA)}$ und β^* sind im allgemeinen nicht unabhängig, es gilt jedoch

$$\text{cov}[\lambda^{(GA)}, \beta^*] = 0 \quad .$$

BEWEIS: Wie von Satz 3.2.2.2.

3.3.3. ASYMPTOTISCHE EIGENSCHAFTEN

Wir führen zunächst eine Zerlegung der Einheitsmatrix ein, die analog zu der in 3.2.3. betrachteten ist. Es ist

$$X' \Sigma_o^{-1} X = \sum_{i=0}^{n-1} X' \left(\frac{1}{\lambda_i^{(o)}} K_i \right) X$$

und folglich

$$I = \sum_{i=0}^{n-1} C_i^{(o)}$$

mit

$$C_i^{(o)} := (X' \Sigma_o^{-1} X)^{-1/2} X' \left(\frac{1}{\lambda_i^{(o)}} K_i \right) X (X' \Sigma_o^{-1} X)^{-1/2}$$

SATZ 3.3.3.1. :

- (1) $0 \leq \text{tr } C_s^{(o)} \leq k$
- (2) $0 \leq \text{tr } [C_s^{(o)}]^2 \leq \text{tr } C_s^{(o)}$
- (3) $F = (\text{tr } C_s^{(o)} C_t^{(o)})$ ist positiv semidefinit.
- (4) $0 \leq \text{tr } C_s^{(o)} C_t^{(o)} \leq \min (\text{tr } C_s^{(o)}, \text{tr } C_t^{(o)})$

BEWEIS :

(1) klar

(2) Es ist

$$C_s^{(o)} = [(X' \Sigma_o^{-1} X)^{-1/2} X' \Sigma_o^{-1/2}] K_s [\Sigma_o^{-1/2} X (X' \Sigma_o^{-1} X)^{-1/2}]$$

Da die Spalten von

$$\Sigma^{-1/2} X (X' \Sigma_o^{-1} X)^{-1/2}$$

orthogonal sind, kann wieder POINCARÉ's Einschließungssatz angewendet werden und man findet für die Eigenwerte $\mu(C_s^{(o)})$ von $C_s^{(o)}$, daß

$$0 \leq \mu(C_s^{(o)}) \leq 1,$$

woraus die Behauptung folgt.

(3), (4) Beweis wie unter 3.2.3.1.

SATZ 3.3.3.2. : Sei $\kappa_h(\lambda_s^{(GA)})$ die h'te Kumulante von $\lambda_s^{(GA)}$. Dann gilt

$$\kappa_h(\lambda_s^{(GA)}) = (h-1)! 2^{h-1} \frac{\lambda_s^h}{r_s^{h-1}} + O\left(\frac{1}{r_s^h}\right)$$

BEWEIS : Wie bei Satz 3.2.3.2. Anstelle der dort auftretenden Matrizen C_i ist nur $\lambda_i^{(o)} C_i^{(o)}$ zu setzen.

SATZ 3.3.3.3. : Es gilt

- (1) $\lambda_s^{(GA)}$ ist asymptotisch unverzerrt
- (2) $\lambda_s^{(GA)}$ ist eine konsistente Schätzfunktion für λ_s
- (3) $\sqrt{r_s} \lambda_s^{(GA)}$ und $\sqrt{r_t} \lambda_t^{(GA)}$ sind asymptotisch unabhängig für $s \neq t$
- (4) $\sqrt{r_s}(\lambda_s^{(GA)} - \lambda_s)$ ist asymptotisch normalverteilt mit Erwartungswert 0 und Varianz $2\lambda_s^2$.

BEWEIS : Wie bei Satz 3.2.3.3.

Satz 3.3.3.3. zeigt, daß in "großen Stichproben" der Einfluß von unterschiedlichen a-priori Annahmen über die Eigenwerte von E verschwindet.

3.4. LOKAL BESTE SCHÄTZER

3.4.1. EINFÜHRUNG

Für den Erwartungswert der ASR-Schätzfunktionen $\lambda^{(A)}$ hatten wir

$$E[\lambda^{(A)}] = V\lambda$$

also

$$E[\lambda_s^{(A)}] = \sum_{t=0}^{n-1} \frac{\text{tr } K_s Q K_t Q}{r_s} \lambda_t .$$

Es liegt nahe, sofern V nichtsingulär ist, die ASR-Schätzfunktion einer Bias-Korrektur zu unterwerfen und die Schätzfunktion

$$\lambda^{(M)} = V^{-1} \lambda^{(A)} ,$$

die dann unverzerrt ist, zu betrachten. Eine hinreichende Bedingung für die Existenz von $\lambda^{(M)}$ ist, daß k (die Anzahl der Regressoren) kleiner ist als $1/2 \min(r_0, r_1, \dots, r_{n-1})$. Die Bedingung ist nicht scharf, jedoch in der Praxis der Sozialforschung stets erfüllt.

SATZ 3.4.1.1. : Sofern $k < 1/2 \min(r_0, r_1, \dots, r_{n-1})$, ist V nichtsingulär.

BEWEIS : Da alle $r_s > 0$ sind, ist V genau dann nichtsingulär, wenn

$$W = (\text{tr } K_s Q K_t Q) = (w_{st})$$

nichtsingulär ist. Definieren wir

$$C_s := (X'X)^{-1/2} X'K_s X (X'X)^{-1/2}$$

so ist

$$w_{ss} = r_s - 2 \text{tr } C_s + \text{tr } C_s^2$$

$$w_{st} = \text{tr } C_s C_t \quad (t \neq s) .$$

Die Matrix W läßt sich also in folgender Form darstellen

$$W = \begin{bmatrix} r_0 - 2\text{tr} C_0 & & & \\ & r_1 - 2\text{tr} C_1 & & \\ & & \ddots & \\ & & & r_{n-1} - 2\text{tr} C_{n-1} \end{bmatrix} + \begin{bmatrix} \text{tr} C_0^2 & \text{tr} C_0 C_1 & \dots & \\ \text{tr} C_1 C_0 & \text{tr} C_1^2 & & \\ \vdots & \vdots & \ddots & \\ \vdots & \vdots & & \text{tr} C_{n-1}^2 \end{bmatrix}$$

sagen wir

$$W = W_1 + W_2 \quad .$$

Für jeden Vektor $x \in \mathbb{R}^n$ gilt folglich

$$x' W x = x' W_1 x + x' W_2 x \geq x' W_1 x \quad ,$$

da nach Satz 3.2.3.1. die Matrix W_2 positiv-definit ist.

Nun ist jedoch

$$x' W_1 x = \sum_{i=0}^{n-1} x_i^2 (r_i - 2\text{tr} C_i)$$

für alle x genau dann positiv, wenn alle Terme

$$r_i - 2 \text{tr} C_i \geq 0$$

und mindestens ein Term

$$r_i - 2 \text{tr} C_i > 0 \quad .$$

Wegen

$$\text{tr} C_i \leq k$$

und

$$\sum \text{tr} C_i = k$$

ist dies sicher dann der Fall, wenn

$$k < 1/2 \min (r_0, r_1 \dots r_{n-1}) \quad .$$

Für den Erwartungswert eines generalisierten ASR-Schätzers $\lambda^{(GA)}$ hatten wir

$$E[\lambda^{(GA)}] = V_0 \lambda$$

also

$$E[\lambda_s^{(GA)}] = \sum_{t=0}^{n-1} \left[\frac{\text{tr } Q_0' K_s Q_0 K_t}{\text{tr } K_s} \right] \lambda_t \quad .$$

Wir betrachten dann die Schätzfunktion

$$\lambda^{(GM)} = V_0^{-1} \lambda^{(GA)} \quad .$$

SATZ 3.4.1.2. : Sofern $k < 1/2 \min (r_0, r_1 \dots r_{n-1})$,
ist V_0 nichtsingulär.

BEWEIS : Definieren wir

$$F := (f_{st}) = (\text{tr } C_s^{(o)} C_t^{(o)})$$

$$D_L := \text{diag}[\lambda_0^{(o)} \dots \lambda_{n-1}^{(o)}]$$

$$D_T := \text{diag}[\text{tr } C_0^{(o)}, \text{tr } C_1^{(o)} \dots \text{tr } C_{n-1}^{(o)}]$$

$$D_R := \text{diag}[r_0, r_1 \dots r_{n-1}]$$

Dann ist

$$V_0 = D_R^{-1} (D_R - 2D_T) + D_R^{-1} D_L F D_L^{-1}$$

genau dann nichtsingulär, wenn

$$W = (D_R - 2D_T) + D_L F D_L^{-1}$$

nichtsingulär ist.

Wir betrachten

$$WD_L^2 = D_R D_L^2 - 2 D_T D_L^2 + D_L F D_L, \quad ,$$

für jedes $x \in \mathbb{R}^n$, $x \neq 0$, ist

$$\begin{aligned} x' WD_L^2 x &= x' (D_R D_L^2 - 2 D_T D_L^2) x + x' D_L F D_L x \\ &\geq x' (D_R D_L^2 - 2 D_T D_L^2) x \end{aligned}$$

da mit F auch $D_L F D_L$ positiv-semidefinit sein muß.

Nun ist

$$\begin{aligned} x' (D_R D_L^2 - 2 D_T D_L^2) x &= \sum_{i=0}^{n-1} x_i^2 [r_i (\lambda_i^{(0)})^2 - 2 (\text{tr} C_i^{(0)}) (\lambda_i^{(0)})^2] \\ &= \sum_{i=0}^{n-1} (x_i \lambda_i^{(0)})^2 (r_i - 2 \text{tr} C_i) \end{aligned}$$

sicher dann größer 0, wenn

$$r_i > 2 \text{tr} C_i^{(0)}$$

für alle $i = 0, \dots, n-1$.

Mit WD_L^2 ist dann auch W und $D_R^{-1} W$ positiv-definit.

3.4.2. FINITE EIGENSCHAFTEN

Für die Schätzfunktionen $\lambda^{(M)}$ und $\lambda^{(GM)}$ sprechen mehr als heuristische Erwägungen. Es handelt sich um lokal varianzminimale Schätzer in der Klasse der unverzerrten, translationsinvarianten quadratischen Schätzfunktionen für λ .

Wir geben zunächst hinreichende und notwendige Bedingungen für Unverzerrtheit und Translationsinvarianz einer quadratischen Schätzfunktion für λ_s .

SATZ 3.4.2.1. : Eine quadratische Schätzfunktion $y' Ay$ für λ_s , wobei A eine symmetrische Matrix bekannter Koeffizienten ist, ist genau dann translationsinvariant ($y' Ay = (y - Xb)' A (y - Xb)$ für alle $b \in \mathbb{R}^k$) und unverzerrt ($E[y' Ay] = \lambda_s$) wenn (RAO [1973, S. 303])

$$(1) \quad AX = 0$$

$$(2) \quad \text{tr } AK_j = \begin{cases} 1 & \text{für } s = j \\ 0 & \text{für } s \neq j \end{cases}$$

BEWEIS : Der Erwartungswert von $y' Ay$ ist

$$E[y' Ay] = \beta' X' A X \beta + \text{tr } A \Sigma \\ \beta' X' A X \beta + \sum_{j=0}^{n-1} (\text{tr } AK_j) \lambda_j .$$

Eine notwendige und hinreichende Bedingung für Unverzerrtheit ist folglich

$$(1) \quad X' A X = 0$$

$$(2) \quad \text{tr } AK_j = \begin{cases} 1 & \text{für } s = j \\ 0 & \text{für } s \neq j \end{cases} .$$

Da $y' Ay$ translationsinvariant sein soll, muß gelten

$$(y - Xb)' A (y - Xb) = y' Ay$$

also

$$b' X' A X b - 2 y' A X b = 0 ,$$

eine Bedingung, die genau dann erfüllt ist, wenn

$$AX = 0 .$$

Die Varianz einer quadratischen, unverzerrten, translationsinvarianten Schätzfunktion für λ_s hängt in solcher Weise von

den unbekannten Parametern λ ab, daß es im allgemeinen nicht möglich sein wird, eine global (für alle möglichen Parameterkonstellationen λ) varianzminimale Schätzfunktion in dieser Klasse von Schätzern zu finden. Wir sind jedoch in der Lage, uns Schätzfunktionen zu verschaffen, die lokale Optimalitätseigenschaften besitzen, die also in einem bestimmten Parameterpunkt $\lambda^{(0)}$ minimale Varianz unter allen Schätzern der angegebenen Klasse aufweisen.

Die Theorie derartiger Schätzfunktionen wurde - unabhängig voneinander - von LAMOTTE [1973] und RAO [1972] entwickelt; RAO bezeichnet diese Schätzer als MINQU (Minimum Norm Quadratic Unbiased) - Schätzer (Anm. 1).

Die Varianz an der Stelle $\lambda^{(0)}$ eines translationsinvarianten, unverzerrten Schätzers $y'A_s y$ für λ_s ist (Satz 3.2.1.1. (2)) proportional zu

$$\text{tr } A_s \Sigma_0 A_s \Sigma_0$$

wobei

$$\Sigma_0 = \lambda_0^{(0)} K_0 + \lambda_1^{(0)} K_1 + \dots + \lambda_{n-1}^{(0)} K_{n-1}.$$

Im folgenden Satz wird die Matrix A_s eines lokal varianzminimalen Schätzers zunächst für $\Sigma_0 = I$, sodann für allgemeines Σ_0 bestimmt.

SATZ 3.4.2.2. :

- (1) Ein Minimum von $\text{tr } A_s^2$, wobei $y'A_s y$ translationsinvariant und unverzerrt für λ_s ist, wird erreicht

Anm. 1: Die übliche Begründung für die Einführung von MINQU-Schätzern benutzt keine Normalitätsannahmen, ist jedoch sehr heuristisch. Unter Annahme einer multivariaten Normalverteilung für y fallen die Begriffe "lokal beste quadratische unverzerrte translationsinvariante Schätzfunktion" und "MINQU-Schätzer" zusammen.

für

$$(*) \quad A_s = \sum_{i=0}^{n-1} \tau_{is} Q K_i Q$$

wobei die τ_{is} Lösungen des Gleichungssystems

$$(**) \quad \sum_{i=0}^{n-1} \tau_{is} (\text{tr } Q K_i Q K_j) = \begin{cases} 1 & \text{für } j = s \\ 0 & \text{sonst} \end{cases}$$

sind. Dabei ist

$$Q := I - X(X'X)^{-1}X'.$$

- (2) Sei $\Sigma_0 = \lambda_0^{(0)} K_0 + \lambda_1^{(0)} K_1 + \dots + \lambda_{n-1}^{(0)} K_{n-1}$. Ein Minimum von $\text{tr } A_s \Sigma_0 A_s \Sigma_0$, wobei $y' A_s y$ translationsinvariant und unverzerrt für λ_s ist, wird erreicht bei

$$(***) \quad A_s = \sum_{i=0}^{n-1} \tau_{is} Q' K_i Q_0$$

wobei die τ_{is} Lösungen des Gleichungssystems

$$(****) \quad \sum_{i=0}^{n-1} \tau_{is} (\text{tr } Q_0' K_i Q_0 K_j) = \begin{cases} 1 & \text{für } j = s \\ 0 & \text{sonst} \end{cases}$$

sind. Dabei ist

$$Q_0 := I - X(X' \Sigma_0^{-1} X)^{-1} X' \Sigma_0^{-1}.$$

BEWEIS:

- (1) (Anm. 1) Translationsinvarianz folgt aus $QX = 0$.
Unverzerrtheit wird durch (**) gesichert.
Sei nun

$$B = A_s + D$$

Anm.: Der Beweis folgt RAO [1973, S.65]

irgendeine symmetrische Matrix, für die $y'By$ translationsinvariant und unverzerrt ist.

Es muß dann gelten, daß auch

$$DX = 0$$

und

$$\text{tr } DK_i = 0 \quad \text{für alle } i = 0, 1, \dots, n-1.$$

Man findet

$$\begin{aligned} \text{tr } B^2 &= \text{tr } (A_s + D)(A_s + D) \\ &= \text{tr } A_s^2 + \text{tr } D^2 + 2 \text{tr } A_s D. \end{aligned}$$

Betrachten wir $\text{tr } A_s D$:

$$\begin{aligned} \text{tr } A_s D &= \text{tr } \sum_{i=0}^{n-1} \tau_{is} QK_i QD \\ &= \sum_{i=0}^{n-1} \tau_{is} \text{tr } QK_i QD \\ &= \sum_{i=0}^{n-1} \tau_{is} \text{tr } K_i QDQ \\ &= 0 \end{aligned}$$

da wegen $DX = 0$ auch $DQ = D$, $QD = D$ und nach Voraussetzung $\text{tr } DK_i = 0$ ist. .

Die Matrix D^2 ist positiv-semidefinit, also ist

$$\text{tr } B^2 = \text{tr } A_s^2 + \text{tr } D^2 \geq \text{tr } A_s^2.$$

(2) Der Beweis läuft ähnlich wie bei (1).

Zunächst ist offenbar

$$Q_0 X = 0$$

und

$$\text{tr } A_s K_j = \begin{cases} 1 & \text{für } s = j \\ 0 & \text{sonst} \end{cases}$$

wegen (****) .

Sei nun

$$B = A_s + D$$

irgendeine symmetrische Matrix, so daß $y'By$ unverzerrt und translationsinvariant ist.

Es muß dann gelten, daß auch $DX = 0$ und

$$\text{tr } DK_i = 0 \quad \text{für alle } i = 0, 1, \dots, n-1 .$$

Nun ist

$$\begin{aligned} \text{tr } B \Sigma_0 B \Sigma_0 &= \text{tr } (A_s + D) \Sigma_0 (A_s + D) \Sigma_0 \\ &= \text{tr } A_s \Sigma_0 A_s \Sigma_0 + 2 \text{tr } A_s \Sigma_0 D \Sigma_0 \\ &\quad + \text{tr } D \Sigma_0 D \Sigma_0 . \end{aligned}$$

Betrachten wir $\text{tr } A_s \Sigma_0 D \Sigma_0$:

$$\begin{aligned} \text{tr } A_s \Sigma_0 D \Sigma_0 &= \sum_{i=0}^{n-1} \tau_{is} \text{tr } Q'_0 K_i Q_0 \Sigma_0 D \Sigma_0 \\ &= \sum_{i=0}^{n-1} \tau_{is} \text{tr } K_i Q_0 \Sigma_0 D \Sigma_0 Q'_0 . \end{aligned}$$

Hier ist

$$\begin{aligned} Q_0 \Sigma_0 D &= (I - X(X' \Sigma_0^{-1} X)^{-1} X' \Sigma_0^{-1}) \Sigma_0 D \\ &= \Sigma_0 D - X(X' \Sigma_0^{-1} X)^{-1} X' D \\ &= \Sigma_0 D \end{aligned}$$

und

$$\begin{aligned} Q \Sigma_0 D \Sigma_0 Q' &= \Sigma_0 D \Sigma_0 Q' \\ &= \Sigma_0 (Q_0 \Sigma_0 D)' \\ &= \Sigma_0 D \Sigma_0 \\ &= \sum_{u=0}^{n-1} \sum_{v=0}^{n-1} \lambda_u \lambda_v K_u D K_v \quad . \end{aligned}$$

Es folgt

$$\text{tr } A_s \Sigma_0 D \Sigma_0 = \sum_{i=0}^{n-1} \sum_{u=0}^{n-1} \sum_{v=0}^{n-1} \tau_{is} \lambda_u \lambda_v \text{tr } (K_i K_u D K_v) = 0 \quad .$$

Da

$$\text{tr } D \Sigma_0 D \Sigma_0 = \text{tr } (\Sigma_0^{1/2} D \Sigma_0^{1/2}) (\Sigma_0^{1/2} D \Sigma_0^{1/2}) \geq 0$$

ist alles bewiesen.

Da die Schätzer $\lambda^{(M)}$ und $\lambda^{(GM)}$ sich als Linearkombinationen von ASR- bzw. GASR-Schätzern $\lambda^{(A)}$ bzw. $\lambda^{(GA)}$ darstellen lassen, kann die Varianz-Kovarianzmatrix dieser Schätzer aus der Varianz-Kovarianzmatrix von $\lambda^{(A)}$ bzw. $\lambda^{(GA)}$ gewonnen werden. Schreiben wir S für diese Matrix, also $s_{ij} = \text{cov}[\lambda_i^{(A)}, \lambda_j^{(A)}] = 2 r_i^{-1} r_j^{-1} \lambda^{(i,j)} \lambda$, so gilt

$$\text{cov}[\lambda_s^{(M)}, \lambda_t^{(M)}] = \sum_{i=0}^{n-1} \sum_{j=0}^{n-1} \tau_{is} \tau_{jt} \text{cov}[\lambda_i^{(A)}, \lambda_j^{(A)}] \quad \text{und damit}$$

$$\text{var}[\lambda_s^{(M)}] = V_0^{-1} S V_0^{-1} \quad , \quad \text{wobei } V_0^{-1} = (\tau_{is}) \quad . \quad \text{Ebenso hat man sofort, da} \quad \text{cov}[\lambda^{(M)}, \beta^*] = V_0^{-1} \text{cov}[\lambda^{(A)}, \beta^*] = 0 \quad .$$

Aufschlußreicher als diese leicht gewonnen Ergebnisse ist die empirische Beobachtung, daß R sich in "großen" Stichproben nur wenig von einer Einheitsmatrix unterscheidet, daß also die Schätzer $\lambda^{(M)}$ den ASR-Schätzern in derartigen Stichproben

"fast" gleich sind. Diese Beobachtung stellt die Grundlage für die asymptotische Theorie der Schätzer $\lambda^{(M)}$ bzw. $\lambda^{(GM)}$ dar.

3.4.3. ASYMPTOTISCHE EIGENSCHAFTEN

$A = (a_{ij})$ sei eine Matrix konstanter Dimension (n, n) . Wir schreiben für eine Abbildung $g : \mathbb{N} \rightarrow \mathbb{R}$.

$$A = O(g(N))$$

wenn alle Elemente der Matrix $(\frac{a_{ij}}{g(N)})$ beim Grenzübergang $N \rightarrow \infty$ gleichmäßig beschränkt bleiben.

Seien nun $\lambda^{(GM)} = V_O^{-1} \lambda^{(GA)}$ die "lokal besten" Schätzer für λ an der Stelle $\lambda^{(o)}$.

SATZ 3.4.3.1. : Es gilt

$$V_O = I + O\left(\frac{1}{N}\right)$$

und für alle hinreichend großen N damit auch

$$V_O^{-1} = I + O\left(\frac{1}{N}\right).$$

BEWEIS : In der Notation von Satz 3.4.1.2. ist

$$\begin{aligned} V_O &= D_R^{-1} (D_R - 2D_T) + D_R^{-1} D_L F D_L^{-1} \\ &= I - 2 D_R^{-1} D_T + D_R^{-1} D_L F D_L^{-1} \end{aligned}$$

Die Matrizen F und D_T , damit auch $D_L F D_L^{-1}$ sind gleichmäßig beschränkt (etwa durch O und k), wie Satz 3.3.3.1. zeigt.

Damit gilt

$$V_0 = I + O\left(\frac{1}{N}\right),$$

denn nach den Voraussetzungen, die unseren asymptotischen Betrachtungen zugrundeliegen, bleibt $\frac{N}{r_i}$ beschränkt für alle $i = 0, 1, \dots, n-1$.

Wählen wir N so groß, daß etwa alle $r_i > 2k$, so existiert V_0^{-1} und auch die Neumann-Entwicklung dieser Matrix

$$\begin{aligned} V_0^{-1} &= (I + O\left(\frac{1}{N}\right))^{-1} \\ &= I - O\left(\frac{1}{N}\right) + [O\left(\frac{1}{N}\right)]^2 + \dots \\ &= I + O\left(\frac{1}{N}\right) \end{aligned}$$

$\lambda^{(GA)}$ und $\lambda^{(GM)}$ sind also asymptotisch äquivalente Schätzfunktionen, sie unterscheiden sich nur um einen Term der Größenordnung $O\left(\frac{1}{N}\right)$. Damit ist Konsistenz von $\lambda^{(GM)}$ sofort klar - es gilt weiter, daß $\lambda^{(GM)}$ die gleiche asymptotische Verteilung wie $\lambda^{(GA)}$ besitzt:

SATZ 3.4.3.2. : $\sqrt{r_s}(\lambda_s^{(GM)} - \lambda_s)$ ist asymptotisch normalverteilt mit Erwartungswert 0 und Varianz $2\lambda_s^2$.

BEWEIS : Es ist

$$\begin{aligned} \sqrt{r_s}(\lambda_s^{(GM)} - \lambda_s) &= \sqrt{r_s}(\lambda_s^{(GA)} + O\left(\frac{1}{N}\right)\lambda_s^{(GA)} - \lambda_s) \\ &= \sqrt{r_s}(\lambda_s^{(GA)} - \lambda_s) + O\left(\frac{\sqrt{r_s}}{N}\right)\lambda_s^{(GA)}. \end{aligned}$$

$\lambda_s^{(GA)}$ konvergiert stochastisch gegen λ , folglich konvergiert $O\left(\frac{\sqrt{r_s}}{N}\right)\lambda_s^{(GA)}$ stochastisch gegen 0.

Es besitzt also

$$\sqrt{r_s}(\lambda_s^{(GM)} - \lambda_s)$$

die gleiche Grenzverteilung wie

$$\sqrt{r_s}(\lambda_s^{(GA)} - \lambda_s) ,$$

diese war nach Satz 3.3.3.3. gleich der hier behaupteten Normalverteilung.

3.5. BALESTRA-SCHÄTZER

3.5.1. EINFÜHRUNG

BALESTRA [1973] suchte in einem Modell der Form

$$y \sim \mathcal{N}[X\beta, \lambda_0 K_0 + \lambda_1 K_1]$$

mit $K_0^2 = K_0$, $K_1^2 = K_1$, $K_0 K_1 = K_1 K_0 = 0$ nachzuweisen, daß Schätzfunktionen des Typs

$$\lambda_s^{(B)} = y' \left[\frac{K_s - K_s X (X' K_s X)^{-1} X' K_s}{\text{tr}(K_s) - k} \right] y \quad (s=0,1)$$

in der Klasse der unverzerrten quadratischen Schätzfunktionen minimale Varianz gleichmäßig im gesamten Parameterraum besitzen. Wir werden im folgenden zeigen, daß diese Behauptung falsch ist.

Die von BALESTRA vorgeschlagenen Schätzfunktionen weisen darüber hinaus einen (unnötigen) Defekt auf - sie setzen nämlich Nicht-singularität aller Matrizen $X' K_s X$ voraus. Diese Voraussetzung

ist in unserem Modell bereits dann nicht erfüllt, wenn ein konstanter Term geschätzt werden soll (also X die Spalte $\mathbb{1}$ enthält), da dann $\mathbb{1}'(I - \Delta_1)\mathbb{1} = 0$.

Wir betrachten anstelle der von BALESTRA vorgeschlagenen Schätzer daher die modifizierten Schätzfunktionen

$$\lambda_s^{(B)} = y' \left[\frac{K_s - K_s X (X' K_s X)^{-1} X' K_s}{\text{tr } K_s - \text{rg } X' K_s X} \right] y \quad (s=0, 1, \dots, n-1)$$

wobei $(X' K_s X)^{-}$ irgendeine generalisierte Inverse von $(X' K_s X)$ darstellt. (Anm. 1)

Da die Einführung von $\lambda_s^{(B)}$ als "beste" quadratische unverzerrte Schätzfunktion unkorrekt ist, seien diese Schätzfunktionen durch ein heuristisches Argument motiviert.

Transformiert man die Kovarianzmatrix des Modells

$$y = X\beta + u$$

auf Diagonalform, betrachtet also

$$T'y = T'X\beta + T'u$$

mit

$$T'u = [0, \lambda_0 I_{r_0} \oplus \lambda_1 I_{r_1} \oplus \dots \oplus \lambda_{n-1} I_{r_{n-1}}]$$

$$T = [T_0 | T_1 | \dots | T_{n-1}]$$

$$T_i T_i' = K_i, T_i' T_i = I \quad ,$$

so zerfällt das Modell in n unabhängige Partialmodelle

$$T_i' y = T_i' X \beta + T_i' u$$

mit

$$T_i' u \sim \mathcal{W}[0, \lambda_i I_{r_i}] \quad .$$

Anm. 1: Für beliebige Matrix Z heißt Z^- generalisierte Inverse von Z , wenn die Beziehung $ZZ^-Z = Z$ erfüllt ist.

Im i 'ten Partialmodell ist β im allgemeinen nicht eindeutig schätzbar, da

$$\begin{aligned}\hat{\beta}_{(i)} &= (X'T_i T_i' X)^{-1} X'T_i T_i' y \\ &= (X'K_i X)^{-1} X'K_i y\end{aligned}$$

bei Singularität von $(X'K_i X)$ mit der Wahl einer speziellen generalisierten Inverse variiert; wohl aber bleiben die Residuen

$$\begin{aligned}\epsilon_i &= T_i' y - T_i' X \hat{\beta}_{(i)} \\ &= (T_i' - T_i' X (X' T_i T_i' X)^{-1} X' T_i T_i') y\end{aligned}$$

und damit auch die übliche Schätzung für die Varianz λ_i der Störungen:

$$\lambda_i^{(B)} = \frac{\epsilon_i' \epsilon_i}{\text{tr}(K_i) - \text{rg}(X'K_i X)} = y' \left[\frac{K_i - K_i X (X'K_i X)^{-1} X'K_i}{\text{tr}(K_i) - \text{rg}(X'K_i X)} \right] y$$

bei unterschiedlicher Wahl von $(X'K_i X)^{-}$ invariant.

3.5.2. FINITE EIGENSCHAFTEN

SATZ 3.5.2.1. : Für die BALESTRA-Schätzer $\lambda_s^{(B)}$ gilt

(1) $\lambda_s^{(B)}$ existiert und ist nichtnegativ für

$$\text{tr}(K_s) > \text{rg}(X'K_s X)$$

(2) $\lambda_s^{(B)}$ ist translationsinvariant

(3) Die Zufallsvariable

$$d_s = \frac{\lambda_s^{(B)}}{\lambda_s}$$

ist zentral chiquadratverteilt mit

$$ds := \text{tr}(K_S) - \text{rg}(X'K_S X)$$

Freiheitsgraden.

$$(4) \quad E[\lambda_s^{(B)}] = \lambda_s$$

$$(5) \quad \text{var}[\lambda_s^{(B)}] = 2 d_s^{-1} \lambda_s^2$$

$$(6) \quad \lambda_s^{(B)} \quad \text{und} \quad \lambda_t^{(B)} \quad \text{sind unabhängig sofern} \quad s \neq t$$

$$(7) \quad \text{Seien} \quad \lambda_0^{(o)}, \lambda_1^{(o)} \dots \lambda_{n-1}^{(o)} \quad \text{positive Konstante,} \\ \Sigma_0 := \lambda_0^{(o)} K_0 + \dots + \lambda_{n-1}^{(o)} K_{n-1} \quad , \text{ dann sind} \\ \lambda_s^{(B)} \quad \text{und} \quad \hat{\beta} = (X' \Sigma_0^{-1} X)^{-1} X' \Sigma_0^{-1} y \quad \text{unabhängig.}$$

BEWEIS:

(1) Die Matrix $K_S - K_S X (X' K_S X)^{-1} X' K_S$ ist für beliebige generalisierte Inverse $(X' K_S X)^{-}$ von $X' K_S X$ idempotent, also auch positiv-semidefinit.

(2) Wegen

$$(X' K_S X) (X' K_S X)^{-} (X' K_S X) = (X' K_S X)$$

hat man für beliebiges $b \in \mathbb{R}^k$ sofort

$$(y - Xb)' (K_S - K_S X (X' K_S X)^{-} X' K_S) (y - Xb) \\ = \lambda_s^{(B)} - 2 y' [K_S X - K_S X (X' K_S X)^{-} X' K_S X] b \quad .$$

Zu zeigen bleibt, daß der Term auf der rechten Seite verschwindet.

Wir benutzen dazu eine explizite Darstellung der generalisierten Inversen einer Matrix Z vom Typ (a, b) . Jede Matrix Z vom Rang c besitzt die

Singulärwertzerlegung

$$Z = L' \begin{bmatrix} \Lambda & 0 \\ 0 & 0 \end{bmatrix} R$$

mit $\Lambda \in \mathbb{R}^{c,c}$ Diagonalmatrix mit positiven Diagonalelementen, sowie L und R orthogonal vom Typ (a,a) bzw. (b,b) .

Jede generalisierte Inverse Z^- von Z läßt sich dann in der Form

$$Z^- = R' \begin{bmatrix} \Lambda^{-1} & S_{10} \\ S_{01} & S_{11} \end{bmatrix} L$$

mit beliebigen $(c,a-c)$, $(b-c,c)$, $(b-c,a-c)$ -Matrizen S_{10} , S_{01} , S_{11} darstellen. Man überzeugt sich nämlich durch Rechnung sofort, daß

$$ZZ^-Z = Z,$$

andererseits kann jede Matrix Z^- in der Form

$$Z^- = R'SL$$

geschrieben werden (man wähle nur $S = RZ^-L'$), so daß mit

$$S = \begin{bmatrix} S_{00} & S_{10} \\ S_{01} & S_{11} \end{bmatrix}$$

der Ansatz $ZZ^-Z = Z$ auf

$$L' \begin{bmatrix} \Lambda & 0 \\ 0 & 0 \end{bmatrix} \begin{bmatrix} S_{00} & S_{10} \\ S_{01} & S_{11} \end{bmatrix} \begin{bmatrix} \Lambda & 0 \\ 0 & 0 \end{bmatrix} L = L' \begin{bmatrix} \Lambda & 0 \\ 0 & 0 \end{bmatrix} L$$

und

$$L' \begin{bmatrix} \Lambda S_{00} \Lambda & 0 \\ 0 & 0 \end{bmatrix} L = L' \begin{bmatrix} \Lambda & 0 \\ 0 & 0 \end{bmatrix} L$$

führt, woraus $S_{00} = \Lambda^{-1}$ folgt.

Wenn also die Singulärwertzerlegung von $K_S X$

$$K_S X = L' \begin{bmatrix} \Lambda & 0 \\ 0 & 0 \end{bmatrix} Q$$

lautet, so wird

$$\begin{aligned} K_S X (X' K_S X)^{-1} X' K_S X &= L' \begin{bmatrix} \Lambda & 0 \\ 0 & 0 \end{bmatrix} \begin{bmatrix} \Lambda^{-2} & S_{01} \\ S_{10} & S_{11} \end{bmatrix} \begin{bmatrix} \Lambda^2 & 0 \\ 0 & 0 \end{bmatrix} R \\ &= L' \begin{bmatrix} \Lambda & 0 \\ 0 & 0 \end{bmatrix} R = K_S X \end{aligned}$$

(3) Es ist

$$\frac{(K_S - K_S X (X' K_S X)^{-1} X' K_S) \Sigma}{\lambda_s} = K_S - K_S X (X' K_S X)^{-1} X' K_S$$

idempotent mit

$$\begin{aligned} \text{rg}(K_S - K_S X (X' K_S X)^{-1} X' K_S) &= \text{tr}(K_S - K_S X (X' K_S X)^{-1} X' K_S) \\ &= \text{tr}(K_S) - \text{tr}[(X' K_S X)^{-1} X' K_S X] \\ &= \text{tr}(K_S) - \text{tr} R' \begin{bmatrix} \Lambda^{-2} & S_{01} \\ S_{10} & S_{11} \end{bmatrix} \begin{bmatrix} \Lambda^2 & 0 \\ 0 & 0 \end{bmatrix} R \\ &= \text{tr}(K_S) - \text{tr} \begin{bmatrix} I & 0 \\ S_{10} \Lambda^2 & 0 \end{bmatrix} \\ &= \text{tr} K_S - \text{rg} X' K_S X \\ &= d_s \end{aligned}$$

(4), (5) folgen direkt aus (3)

(6) ist klar wegen $K_s \Sigma K_t = 0$ für $s \neq t$.

(7) Zu zeigen ist

$$[X' \Sigma_0^{-1} X]^{-1} X' \Sigma_0^{-1} \Sigma [K_S - K_S X (X' K_S X)^{-1} X' K_S] = 0 ,$$

was sich durch Rechnung direkt ergibt, wenn man die Beziehung

$$K_S X (X' K_S X)^{-1} X' K_S X = K_S X$$

benutzt.

BALESTRA-Schätzer haben also recht angenehme Eigenschaften (insbesondere die Chiquadratverteilung) - die Behauptung von BALESTRA, daß es sich um beste unverzerrte quadratische Schätzfunktionen handelt, ist jedoch falsch. Es genügt, dies am einfachsten Beispiel nachzuweisen:

SATZ 3.5.2.2. : Sei $\lambda_0^{(M)}$ der "MINQU"-Schätzer für λ_0 in einem Zweiebenenmodell, $\lambda_0^{(B)}$ der BALESTRA-Schätzer für λ_0 , dann gilt

$$\text{var}[\lambda_0^{(M)}] < \text{var}[\lambda_0^{(B)}]$$

falls $r_1 = \text{tr}(K_1) > k$, wobei die Varianz an der Stelle $\lambda_0 = \lambda_1 = 1$ berechnet wird.

BEWEIS : $\lambda_0^{(M)}$ ist eine gewichtete Summe

$$\lambda_0^{(M)} = \alpha_0 r_0 \lambda_0^{(A)} + \alpha_1 r_1 \lambda_1^{(A)}$$

der ASR-Schätzer für λ_0 und λ_1 .

Damit $\lambda_0^{(M)}$ unverzerrt ist, muß gelten

$$\alpha_0 = \frac{t_1}{t_0 t_1 - c_{01}^2}$$

$$\alpha_1 = - \frac{c_{01}}{t_0 t_1 - c_{01}^2} .$$

Hierbei wurde

$$c_o := \text{tr } C_o = \text{tr } (X'X)^{-1/2} X' K_o X (X'X)^{-1/2}$$

$$c_1 := \text{tr } C_1 = \text{tr } (X'X)^{-1/2} X' K_1 X (X'X)^{-1/2}$$

$$c_{o1} := \text{tr } C_o C_1$$

$$c_{oo} := \text{tr } C_o^2$$

$$c_{11} := \text{tr } C_1^2$$

$$t_o := r_o + c_{oo} - 2 c_o$$

$$t_1 := r_1 + c_{11} - 2 c_1$$

gesetzt.

Betrachtet man Varianzen an der Stelle $\lambda_o = \lambda_1 = 1$,
so hat man nach längeren Umformungen:

$$\begin{aligned} \text{var}[\lambda_o^{(M)}] &= \frac{1}{(t_o t_1 - c_{o1}^2)^2} [r_o^2 t_o^2 \text{var}[\lambda_o^{(A)}] + r_1^2 c_{o1}^2 \text{var}[\lambda_1^{(A)}] \\ &\quad - 2 r_o t_1 r_1 c_{o1} \text{cov}[\lambda_o^{(A)}, \lambda_1^{(A)}]] \\ &= \frac{2 t_1}{t_o t_1 - c_{o1}^2} \\ &= \frac{2}{r_o - [\frac{c_{o1}^2}{t_1} + 2c_o - c_{oo}]} \end{aligned}$$

wenn man

$$\text{var}[\lambda_o^{(A)}] = \frac{2t_o}{r_o^2}$$

$$\text{var}[\lambda_1^{(A)}] = \frac{2t_1}{r_1^2}$$

$$\text{cov}[\lambda_0^{(A)}, \lambda_1^{(A)}] = \frac{2c_{01}}{r_0 r_1}$$

benutzt.

Sei p.B.d.A. $\text{rg}(X'K_0X) = k$, dann ist

$$\text{var}[\lambda_0^{(B)}] = \frac{2}{r_0 - k}.$$

Es bleibt also zu zeigen, daß

$$\frac{c_{01}^2}{t_1} + 2 c_0 - c_{00} < k.$$

Wir wissen nun, daß mit $r_1 > k$ auch

$$r_1 c_{11} > c_{11} (c_{00} + 2 c_{01} + c_{11})$$

wegen

$$k = c_{00} + 2 c_{01} + c_{11}.$$

Da die Matrix

$$\begin{bmatrix} c_{00} & c_{01} \\ c_{10} & c_{11} \end{bmatrix}$$

positiv-semidefinit ist, gilt

$$c_{00} c_{11} > c_{01}^2$$

also

$$r_1 c_{11} > c_{01}^2 + 2 c_{01} c_{11} + c_{11}^2,$$

woraus

$$c_{01}^2 < r_1 c_{11} - 2 c_{01} c_{11} - c_{11}^2 = c_{11} (r_1 + c_{11} - 2 c_{01})$$

und

$$\frac{c_{01}}{t_1} < c_{11}$$

folgt.

Es gilt also, wie behauptet,

$$\frac{c_{01}^2}{t_1} + 2 c_0 - c_{00} < c_{11} + 2 c_0 - c_{00} = c_{11} + 2 c_{01} - c_{00} = k .$$

3.5.3. ASYMPTOTISCHE EIGENSCHAFTEN

Die asymptotische Theorie ist sehr einfach und folgt direkt aus den Eigenschaften der Chiquadratverteilung - wir halten nur die Ergebnisse fest:

- (1) $\lambda_s^{(B)}$ ist konsistent.
- (2) $\sqrt{r_s}(\lambda_s^{(B)} - \lambda_s)$ ist asymptotisch normalverteilt mit Erwartungswert 0 und Varianz $2\lambda_s^2$.

4. SIMULTANE SCHÄTZUNG VON EIGENWERTEN UND REGRESSIONSKOEFFIZIENTEN

4.1. FEEDBACK VON ASR, MINQU UND BALESTRA SCHÄTZERN

Die gewonnenen Schätzer $\hat{\lambda}$ für die Eigenwerte der Kovarianzmatrix können dazu verwendet werden, um "bessere" Schätzer für die Regressionskoeffizienten β zu erhalten: man benutzt Schätzer vom AITKEN-Typ

$$\hat{\beta} = (X' \hat{\Sigma}^{-1} X)^{-1} X' \hat{\Sigma}^{-1} y$$

in der $\hat{\Sigma}$ eine mittels $\hat{\lambda}$ berechnete Schätzung für die Kovarianzmatrix von y ist

$$\hat{\Sigma} = \sum_{i=0}^{n-1} \hat{\lambda}_i K_i.$$

Diese Vorgehensweise ("feedback" von ASR, MINQU oder BALESTRA-Schätzern) ist plausibel; die Analyse der stochastischen Eigenschaften von $\hat{\beta}$ wird jedoch dadurch erschwert, daß es sich bei $\hat{\Sigma}$ um eine Zufallsvariable handelt. So gehört etwa $\hat{\beta}$ bereits nicht mehr zur Klasse der linearen Schätzfunktionen für β und es wird rasch sehr verwickelt, wenn man auch nur so einfache finite Eigenschaften wie etwa die Unverzerrtheit von $\hat{\beta}$ untersuchen will.

Wir beschränken uns daher auf 2 Sätze, in denen asymptotische Eigenschaften der zweiphasigen Schätzer $\hat{\beta}$ für β ausgesprochen werden:

- (1) $\sqrt{N}(\hat{\beta} - \beta)$ konvergiert stochastisch gegen 0, d.h. $\hat{\beta}$ und der AITKEN-Schätzer $\tilde{\beta} = (X' \Sigma^{-1} X)^{-1} X' \Sigma^{-1} y$ sind asymptotisch äquivalent.

- (2) $\hat{\beta}$ ist asymptotisch effizient, $\sqrt{N}(\hat{\beta} - \beta)$ ist asymptotisch multivariat normalverteilt mit Erwartungswert 0 und Kovarianzmatrix $(X' \Sigma^{-1} X)^{-1}$.

Voraussetzung ist bei beiden Sätzen, neben den in Abschnitt 3.2.3. zusammengestellten asymptotischen Annahmen, die Benutzung von konsistenten Schätzfunktionen $\hat{\lambda}$ für die Schätzung von Σ . Alle betrachteten Schätzer für λ sind also zur Bildung zweiphasiger Schätzfunktionen für β geeignet.

SATZ 4.1.1. : Seien $(\hat{\lambda}_0, \hat{\lambda}_1, \dots, \hat{\lambda}_{n-1})$ konsistente Schätzer für $(\lambda_0, \lambda_1, \dots, \lambda_{n-1})$, dann sind AITKEN-Schätzer $\hat{\beta} = (X' \Sigma^{-1} X)^{-1} X' \Sigma^{-1} y$ und zweiphasige Schätzer $\hat{\beta} = (X' \hat{\Sigma}^{-1} X)^{-1} X' \hat{\Sigma}^{-1} y$ (mit $\hat{\Sigma} = \hat{\lambda}_0 K_0 + \dots + \hat{\lambda}_{n-1} K_{n-1}$) asymptotisch äquivalent, d.h. es gilt

$$\text{plim}_{N \rightarrow \infty} \sqrt{N}(\hat{\beta} - \hat{\beta}) = 0.$$

BEWEIS : Wir setzen zur Abkürzung

$$\begin{aligned} S &:= X' \Sigma^{-1} X & q &:= X' \Sigma^{-1} u \\ \hat{S} &:= X' \hat{\Sigma}^{-1} X & \hat{q} &:= X' \hat{\Sigma}^{-1} u; \end{aligned}$$

es gilt dann

$$\begin{aligned} \hat{\beta} &= \beta + S^{-1} q \\ \hat{\beta} &= \beta + S^{-1} \hat{q}. \end{aligned}$$

Für $N \rightarrow \infty$ findet man zunächst (Anm. 1)

$$(1) \quad \frac{1}{N} S \rightarrow V := \sum_{s=0}^{n-1} \frac{k_s}{\lambda_s} S_s$$

$$(2) \quad \frac{1}{N} \hat{S} \xrightarrow{P} V$$

Anm. 1: $\xrightarrow{P}$ bedeutet stochastische Konvergenz, $\xrightarrow{V}$ Konvergenz in Verteilung, $\rightarrow$ gewöhnliche Konvergenz.

Zur sonstigen Bezeichnung siehe Abschnitt 3.2.3.

$$(3) \quad \frac{1}{N} (\hat{S} - S) \xrightarrow{P} 0$$

$$(4) \quad \left[\frac{1}{N} \hat{S} \right]^{-1} \xrightarrow{P} V^{-1}$$

$$(5) \quad \frac{1}{\sqrt{N}} (\hat{q} - q) \xrightarrow{P} 0$$

$$(6) \quad \sqrt{N} (\hat{\beta} - \beta) \xrightarrow{V} \eta \sim \mathcal{N}[0, V^{-1}]$$

Dabei gilt (1) nach Annahme, für $\frac{1}{N} \hat{S}$ findet man

$$\begin{aligned} \text{plim } \frac{1}{N} \hat{S} &= \text{plim } \left[\frac{1}{N} \sum_{s=0}^{n-1} \frac{1}{\lambda_s} X' K_s X \right] \\ &= \sum_{s=0}^{n-1} \text{plim } \left[\frac{1}{\lambda_s} \left(\frac{1}{N} X' K_s X \right) \right] \\ &= \sum_{s=0}^{n-1} \left(\text{plim } \frac{1}{\lambda_s} \right) \left(\lim \frac{1}{N} X' K_s X \right) \\ &= \sum_{s=0}^{n-1} \left(\frac{1}{\lambda_s} \right) (K_s S_s) = V, \end{aligned}$$

eine nichtsinguläre Matrix.

Aus (2) folgt (3), und mit SLUTZKY's Theorem sowie der Tatsache, daß die Inverse eine stetige Funktion der Matrixelemente ist, auch (4).

Für (5) hat man

$$\frac{1}{\sqrt{N}} (\hat{q} - q) = \frac{1}{\sqrt{N}} \sum_{s=0}^{n-1} \left(\frac{1}{\lambda_s} - \frac{1}{\lambda_s} \right) X' K_s u.$$

Es ist

$$E[X' K_s u] = 0$$

$$\text{var}[X' K_s u] = \lambda_s X' K_s X$$

also

$$E\left[\frac{1}{\sqrt{N}} X' K_s u\right] = 0$$

$$\text{var}\left[\frac{1}{\sqrt{N}} X' K_s u\right] = \frac{1}{N} \lambda_s X' K_s X + \lambda_s k_s S_s \quad .$$

Die Folge der Zufallsvariablen $\frac{1}{\sqrt{N}} X' K_s u$ besitzt beschränkte erste und zweite Momente. Wegen

$$\text{plim} \left(\frac{1}{\lambda_s} - \frac{1}{\lambda_s} \right) = 0$$

ist also auch

$$\text{plim} \frac{1}{\sqrt{N}} (\hat{q} - q) = 0 \quad .$$

(6) ist klar, wegen Normalverteilung von $\hat{\beta}$ und

$$E[\hat{\beta}] = \beta$$

$$\text{var}[\hat{\beta}] = (X' \Sigma^{-1} X)^{-1} \quad ,$$

woraus

$$\text{var}[\sqrt{N}(\hat{\beta} - \beta)] \rightarrow \left(\frac{1}{N} X' \Sigma^{-1} X \right)^{-1} = V^{-1}$$

folgt.

Man findet nun für $(\hat{\beta} - \hat{\beta})$ durch eine Reihe von Umformungen:

$$\begin{aligned} \hat{\beta} - \hat{\beta} &= \hat{S}^{-1} \hat{q} - S^{-1} q \\ &= \hat{S}^{-1} (\hat{q} - q) + (\hat{S}^{-1} - S^{-1}) q \\ &= \hat{S}^{-1} (\hat{q} - q) + \hat{S}^{-1} (q - \hat{S} \hat{S}^{-1} q) \\ &= \hat{S}^{-1} (\hat{q} - q) + \hat{S}^{-1} (S(\hat{\beta} - \beta) - \hat{S}(\hat{\beta} - \beta)) \\ &= \hat{S}^{-1} (\hat{q} - q) + \hat{S}^{-1} (S - \hat{S})(\hat{\beta} - \beta) \end{aligned}$$

Der erste Term auf der rechten Seite der folgenden Gleichung

$$\sqrt{N}(\hat{\beta} - \hat{\beta}) = \left(\frac{1}{N} \hat{S} \right)^{-1} \frac{1}{\sqrt{N}} (\hat{q} - q) + \left(\frac{1}{N} \hat{S} \right)^{-1} \left(\frac{1}{N} S - \frac{1}{N} \hat{S} \right) \sqrt{N}(\hat{\beta} - \beta)$$

konvergiert der Wahrscheinlichkeit nach gegen

$$\text{plim } \left(\frac{1}{N} \hat{S}\right)^{-1} \cdot \text{plim } \frac{1}{\sqrt{N}} (\hat{q} - q) = v \cdot 0 = 0 ,$$

der zweite Term gegen

$$\text{plim } \left(\frac{1}{N} \hat{S}\right)^{-1} \cdot \text{plim } \left(\frac{1}{N} S - \frac{1}{N} \hat{S}\right) \cdot \text{plim } \sqrt{N}(\hat{\beta} - \beta) = v \cdot 0 \cdot n$$

also gleichfalls gegen den Nullvektor.

Als Korollar ergibt sich sofort

SATZ 4.1.2. : $\sqrt{N}(\hat{\beta} - \beta)$ ist asymptotisch multivariat normalverteilt mit Erwartungswertvektor 0 und Kovarianzmatrix

$$v^{-1} = \lim \left(\frac{1}{N} X' \Sigma X \right)^{-1}$$

BEWEIS : Wir haben

$$\sqrt{N}(\hat{\beta} - \beta) = \sqrt{N}(\hat{\beta} - \hat{\beta}) + \sqrt{N}(\hat{\beta} - \beta) .$$

Der erste Term auf der rechten Seite konvergiert stochastisch gegen 0, der zweite Term konvergiert in Verteilung gegen eine $\mathcal{N}[0, v^{-1}]$ -verteilte Zufallsvariable - damit ist auch die Grenzverteilung der Summe eine $\mathcal{N}[0, v^{-1}]$ -Verteilung.

Die asymptotische Effizienz von $\hat{\beta}$ folgt aus dieser Tatsache, sowie dem bekannten Resultat, daß $\hat{\beta}$ asymptotisch effizient ist.

4.2. MAXIMUM-LIKELIHOOD-SCHÄTZER

4.2.0. VORBEMERKUNG

Die Untersuchung der Likelihoodfunktion des Modells ist nicht nur von Interesse, um Maximum-Likelihood-Schätzer zu gewinnen - wir werden mit Hilfe der Likelihoodfunktion Aussagen über minimale und vollständige Suffizienz von Statistiken, Identifizierbarkeit des Modells sowie über CRAMER-RAO-Schranken für die Varianz-Kovarianzmatrix möglicher Schätzer erhalten können.

4.2.1. LIKELIHOODFUNKTION UND VOLLSTÄNDIG SUFFIZIENTE STATISTIKEN

Vernachlässigt man Terme, die von den Parametern $\Theta = (\beta, \lambda)$ des Modells nicht abhängen, so lautet die logarithmierte Likelihoodfunktion des Modells

$$S(\Theta|y) = -\frac{1}{2} \log \det (\lambda_0 K_0 + \lambda_1 K_1 + \dots + \lambda_{n-1} K_{n-1}) - \frac{1}{2} (y - X\beta)' (\lambda_0 K_0 + \dots + \lambda_{n-1} K_{n-1})^{-1} (y - X\beta)$$

Im allgemeinen findet man als minimal-suffiziente Statistik nur die $n(k+1)$ -dimensionale Statistik $(y'K_0y \dots y'K_{n-1}y, X'K_0y \dots X'K_{n-1}y)$. Das Modell besitzt im allgemeinen daher keine vollständig suffiziente Statistik.

Es ist nicht nur von theoretischer Wichtigkeit, notwendige und hinreichende Bedingungen für vollständige Suffizienz der $(n+k)$ -dimensionalen Statistik $(y'K_0y, y'K_1y \dots y'K_{n-1}y, X'y)$ anzugeben, da eine in der Praxis der Mehrebenenanalyse häufig auftauchende Konstellation für den Erwartungswertteil des Modells diese Bedingungen erfüllt.

SATZ 4.2.1.1. : Die folgenden Aussagen sind äquivalent:

- (1) die $(n+k)$ -dimensionale Statistik $(y'K_0y, y'K_1y, \dots, y'K_{n-1}y, X'y)$ ist vollständig suffizient.
- (2) es existieren Matrizen $L_0, L_1, \dots, L_{n-1}$, so daß für $i = 0, 1, \dots, n-1$

$$K_i X = X L_i,$$

d.h. $\mathcal{R}(X)$ ist invariant unter K_i .

BEWEIS:

(1) $\rightarrow$ (2). Eine vollständig suffiziente Statistik ist notwendig auch minimal suffizient. Der übliche LEHMANN-SCHEFFÉ-Ansatz führt, ausgehend von der Loglikelihood

$$S(\theta|y) = \alpha(\theta) - \frac{1}{2} \sum_{i=0}^{n-1} y'K_i y \frac{1}{\lambda_i} + \sum_{i=0}^{n-1} y'K_i X \beta \frac{1}{\lambda_i},$$

in der $\alpha(\theta)$ eine über dem Stichprobenraum konstante Funktion darstellt, zu der minimal suffizienten Statistik

$$(y'K_0y, \dots, y'K_{n-1}y, y'K_0X, \dots, y'K_{n-1}X).$$

Seien y und $y_0 (\neq y)$ zwei Punkte des Stichprobenraums mit $y'K_i y = y_0'K_i y_0$ und $(y-y_0)'X = 0$ ($i=0, 1, \dots, n-1$). Da nach Annahme $(y'K_0y, \dots, y'K_{n-1}y, X'y)$ minimal suffizient ist, muß dann auch

$$\sum_{i=0}^{n-1} (y-y_0)'K_i X \beta \frac{1}{\lambda_i} = 0$$

für alle (β, λ) sein, d.h. aber es muß für $i=0, 1, \dots, n-1$ aus $(y-y_0)'X = 0$ auch $(y-y_0)'K_i X = 0$ folgen. Dies ist offenbar genau dann der Fall, wenn

$$K_i X = X L_i$$

für irgendeine Matrix L_i .

(2) $\rightarrow$ (1). Sei $K_i X = X L_i$ für alle $i=0,1,\dots,n-1$, dann lautet die Loglikelihood

$$\begin{aligned} S(\theta|y) &= \alpha(\theta) - \frac{1}{2} \sum_{i=0}^{n-1} \frac{y' K_i y}{\lambda_i} + \sum_{i=0}^{n-1} \frac{y' K_i X \beta}{\lambda_i} \\ &= \alpha(\theta) - \frac{1}{2} \sum_{i=0}^{n-1} \frac{y' K_i y}{\lambda_i} + y' X \sum_{i=0}^{n-1} \frac{L_i \beta}{\lambda_i} \\ &= \alpha(\theta) - \frac{1}{2} \sum_{i=0}^{n-1} \frac{y' K_i y}{\lambda_i} + y' X (X' X)^{-1} X' \Sigma^{-1} X \beta \end{aligned}$$

da aus der Gleichung

$$K_i X = X L_i$$

folgt, daß

$$L_i = (X' X)^{-1} X' K_i X.$$

Die Abbildung

$$F : \Omega_1 \times \Omega_2 \rightarrow \mathbb{R}^k \times \mathbb{R}^n,$$

die dem Vektor

$$\theta = \begin{bmatrix} \beta \\ \lambda \end{bmatrix}$$

den Vektor

$$F(\theta) = \theta^* = \begin{bmatrix} \sum_{i=0}^{n-1} (X' X)^{-1} X' K_i \frac{1}{\lambda_i} X \\ \lambda \end{bmatrix} = \begin{bmatrix} \theta^* \\ \lambda \end{bmatrix}$$

zuordnet, ist stetig differenzierbar und die Matrix der partiellen Ableitungen

$$J = \frac{\partial \theta^*}{\partial \theta} = \begin{bmatrix} (X'X)^{-1}X'\Sigma^{-1}X & -\frac{1}{\lambda_i^2}(X'X)^{-1}(X'K_iX)\delta_i \\ \hline C & I \end{bmatrix}$$

ist in ganz $\Omega_1 \times \Omega_2$ von vollem Rang $(n+k)$, da die Jacobische Determinante

$$\det J = \det(X'X)^{-1}X'\Sigma^{-1}X \neq 0$$

in ganz $\Omega_1 \times \Omega_2$ nach Definition von Ω_2 .

$F[\Omega_1 \times \Omega_2]$ enthält damit ein offenes $(n+k)$ -dimensionales Intervall; es kann also zusammen mit der Tatsache, daß

$$S(\theta^*|y) = \alpha(\theta^*) + y'X\theta_1^* - \frac{1}{2} \sum_{i=0}^{n-1} (y'K_i y)\lambda_i$$

die Loglikelihoodfunktion einer $(n+k)$ -dimensionalen Exponentialverteilung ist, Theorem 1 (S. 132) von LEHMANN [1959] angewendet werden.

Ob eine derartige vollständig suffiziente Statistik existiert, kann in der Praxis mit Hilfe der folgenden Beziehungen leicht überprüft werden:

SATZ 4.2.1.2. : $(y'K_0 y, \dots, y'K_{n-1} y, X'y)$ ist genau dann vollständig suffizient, wenn für alle $i=0, 1, \dots, n-1$ die Beziehung

$$(*) \quad \text{tr}[(X'X)^{-1}X'K_i X] = \text{tr}[(X'X)^{-1}X'K_i X]^2$$

erfüllt ist.

BEWEIS : Es ist nur zu zeigen, daß $(*)$ äquivalent ist zu $K_i X = X L_i$ für $i=0, 1, \dots, n-1$.

(1) Sei $K_i X = X L_i$, also

$$L_i = (X'X)^{-1}X'K_i X.$$

Aus

$$K_i^2 X = X L_i^2$$

folgt

$$L_i^2 = L_i$$

und damit (*) .

(2) Sei

$$\text{tr}(X'X)^{-1} X' K_i X = \text{tr}[(X'X)^{-1} X' K_i X]^2$$

für alle $i=0,1,\dots,n-1$.

Benutzen wir die Bezeichnung

$$C_i := (X'X)^{-1/2} X' K_i X (X'X)^{-1/2} ,$$

so gilt dann auch

$$\text{tr } C_i^2 = \text{tr } C_i$$

also

$$\text{tr}(C_i - C_i^2) = 0 .$$

Nun ist

$$C_i - C_i^2 = (X'X)^{-1/2} X' K_i [I - X(X'X)^{-1} X'] K_i X (X'X)^{-1/2}$$

positiv semidefinit, also

$$0 = \text{tr}(C_i - C_i^2) \iff C_i = C_i^2 .$$

Aus $C_i = C_i^2$ folgt

$$(X' K_i X)(X'X)^{-1} (X' K_i X) - X' K_i X = 0$$

und es ist mit

$$[X' K_i - (X' K_i X)(X'X)^{-1} X'] [K_i X - X(X'X)^{-1} X' K_i X] = 0$$

auch notwendig

$$K_i X = X(X'X)^{-1}X'K_i X \quad .$$

In Modellen mit Regressoren, die unter allen K_i invariant bleiben, ist die zu Beginn des 3. Kapitels geschilderte Schwierigkeit des hier untersuchten Schätzproblems - die Nichtkommutativität von Projektionen auf den Regressorenraum und auf die Eigenräume von Σ - nicht vorhanden. Bezeichnet P den Projektor $X(X'X)^{-1}X'$ auf den Spaltenraum von X und Q die Matrix $I-P$, so gelten in derartigen Modellen die (leicht nachzuprüfenden) Beziehungen:

$$\begin{aligned} (1) \quad K_i P &= P K_i P & (2) \quad Q K_i X &= 0 \\ (3) \quad P K_i P &= K_i P K_i & (4) \quad K_i P K_j &= 0 \\ (5) \quad Q K_i Q &= K_i Q K_i & (6) \quad Q K_i &= K_i Q \end{aligned}$$

verbal also insbesondere:

- Die Projektionen von X auf die Eigenräume von Σ liegen im Regressionsraum, bzw. sind orthogonal zum Residualraum (Schätzungen für λ_i bleiben also unbeeinflusst von etwa in X auftretenden Mehrebeneneffekten).
- Es ist gleichgültig, ob man zur Schätzung von λ_i die Residuen auf den Eigenraum von K_i projiziert oder die Projektionen von y auf die Eigenräume mittels Q vom Einfluß von X befreit.

Ein für Anwendungen der Mehrebenenanalyse typisches Modell sucht die Beobachtungen y zu erklären durch Effekte ein- und desselben Variablensatzes Z , die auf verschiedenen Ebenen auftreten. Es wird also folgendes Modell für den Erwartungswert von y postuliert:

$$E[y] = Z\beta_0 + \Delta_1 Z\beta_1 + \dots + \Delta_{n-1} Z\beta_{n-1} .$$

Eine Reparametrisierung überführt das Modell in

$$E[y] = K_0 Z\alpha_0 + K_1 Z\alpha_1 + \dots + K_{n-1} Z\alpha_{n-1} = X\alpha$$

wobei

$$\alpha_0 = \beta_0; \alpha_1 - \alpha_0 = \beta_1; \dots; \alpha_{n-1} - \alpha_{n-2} = \beta_{n-1}$$

und

$$X = [K_0 Z | K_1 Z | \dots | K_{n-1} Z] .$$

Für $(X'X)^{-1}X'K_iX$ findet man

$$(X'X)^{-1}X'K_iX = \begin{bmatrix} Z'K_0Z & & & \\ & \ddots & & \\ & & Z'K_iZ & \\ & & & \ddots \\ & & & & Z'K_{n-1}Z \end{bmatrix}^{-1} \begin{bmatrix} 0 & & & \\ & \ddots & & \\ & & Z'K_iZ & \\ & & & \ddots \\ & & & & 0 \end{bmatrix}$$

$$= \begin{bmatrix} 0 & & & \\ & \ddots & & \\ & & I & \\ & & & \ddots \\ & & & & 0 \end{bmatrix} ,$$

es ist also

$$\text{tr}[(X'X)^{-1}X'K_iX]^2 = \text{tr}[(X'X)^{-1}X'K_iX] .$$

Etwas verallgemeinert:

In jedem Modell der Form

$$E[y] = K_0 Z_0 \alpha_0 + K_1 Z_1 \alpha_1 + \dots + K_{n-1} Z_{n-1} \alpha_{n-1} = X\alpha$$

gilt

$$\text{tr}[(X'X)^{-1}X'K_iX]^2 = \text{tr}[(X'X)^{-1}X'K_iX]$$

und existiert somit eine $(n+k)$ -dimensionale vollständig suffiziente Statistik.

Falls die Bedingung von Satz 4.2.1.2. erfüllt ist, sind Maximum-Likelihood-Schätzer $\hat{\beta}, \hat{\lambda}$ (und damit sofort beste unverzerzte Schätzfunktionen) explizit angebar: wir werden bei der Betrachtung der Likelihoodgleichungen finden, daß dann $\hat{\beta}$ gleich β^* (den OLS-Schätzfunktionen) und $\hat{\lambda}$ gleich $\lambda^{(A)}$ (den ASR-Schätzfunktionen) ist.

4.2.2. LIKELIHOODGLEICHUNGEN UND INFORMATIONSMATRIX

Die ersten und zweiten partiellen Ableitungen der Loglikelihoodfunktion lassen sich mit Hilfe von Techniken der Matrizen-differentiation (MCRAE [1975]) relativ bequem ermitteln:

$$\frac{\partial S}{\partial \beta} = X' \Sigma^{-1} y - X' \Sigma^{-1} X \beta = \sum_{i=0}^{n-1} \frac{1}{\lambda_i} X' K_i y - \sum_{i=0}^{n-1} \frac{1}{\lambda_i} X' K_i X \beta$$

$$\begin{aligned} \frac{\partial S}{\partial \lambda_i} &= -\frac{1}{2} \text{tr}(\Sigma^{-1} K_i) + \frac{1}{2} (y - X\beta)' \Sigma^{-1} K_i \Sigma^{-1} (y - X\beta) \\ &= -\frac{r_i}{2\lambda_i} + \frac{(y - X\beta)' K_i (y - X\beta)}{2\lambda_i^2} \end{aligned}$$

$$\frac{\partial^2 S}{\partial \beta \partial \beta'} = -X \Sigma^{-1} X = -\sum_{i=0}^{n-1} \frac{1}{\lambda_i} X' K_i X$$

$$\begin{aligned} \frac{\partial^2 S}{\partial \beta \partial \lambda_i} &= -X' \Sigma^{-1} K_i \Sigma^{-1} y + X' \Sigma^{-1} K_i \Sigma^{-1} X \beta \\ &= -\frac{1}{\lambda_i^2} X' K_i (y - X\beta) \end{aligned}$$

$$\frac{\partial^2 S}{\partial \lambda_i \partial \lambda_j} = 0 \quad (i \neq j)$$

$$\frac{\partial^2 S}{\partial \lambda_i^2} = \frac{r_i}{2\lambda_i^2} - \frac{(y - X\beta)' K_i (y - X\beta)}{\lambda_i^3}$$

Die zweiten Ableitungen liefern die Informationsmatrix

$$J = - E \frac{\partial^2 \hat{S}}{\partial \theta \partial \theta'}$$

$$= \begin{bmatrix} E[X' \Sigma^{-1} X] & E\left[\frac{X' K_i (y - X\beta)}{\lambda_i^2}\right] \\ E\left[\frac{(y - X\beta)' K_i X}{\lambda_i^2}\right] & E\left[\frac{(y - X\beta)' K_i (y - X\beta)}{\lambda_i^2} - \frac{r_i}{2\lambda_i^2}\right] \end{bmatrix}$$

Hier ist

$$E[X' K_i (y - X\beta)] = X' K_i E[y - X\beta] = 0$$

und

$$E[(y - X\beta)' K_i (y - X\beta)] = \lambda_i r_i ,$$

also

$$J = \begin{bmatrix} X' \Sigma^{-1} X & 0 \\ 0 & D \end{bmatrix} \quad \text{mit } D = \text{diag} \left[\frac{1}{2} \frac{r_i}{\lambda_i^2} \right] .$$

Die CRAMÉR-RAO-Schranken lassen sich durch Inversion von J angeben und liefern Ungleichungen für die Varianzen unverzerrter Schätzer $\hat{\beta}$ und $\hat{\lambda}_s$:

$$\text{var}[\hat{\beta}] \geq (X' \Sigma^{-1} X)^{-1} \quad (\text{d.h. } \text{var}[\hat{\beta}] - (X' \Sigma^{-1} X)^{-1} \text{ ist positiv definit})$$

$$\text{var}[\hat{\lambda}_s] \geq \frac{2\lambda_s^2}{r_s} .$$

Mit dieser Tatsache, sowie den asymptotischen Resultaten von Kapitel 3, folgt, daß alle dort betrachteten Schätzer für λ_s BAN-Schätzer (beste asymptotisch normalverteilte Schätzfunktionen) sind.

4.2.3. EIN ALGORITHMUS ZUR LÖSUNG
DER LIKELIHOODGLEICHUNGEN

Schreibt man die Likelihoodgleichungen in der Form:

$$(*) \quad \hat{\beta} = (X' \left(\sum_{s=0}^{n-1} \frac{1}{\hat{\lambda}_s} K_s \right) X)^{-1} X' \left(\sum_{s=0}^{n-1} \frac{1}{\hat{\lambda}_s} K_s y \right)$$

$$(**) \quad \hat{\lambda}_s = \frac{(y - X\hat{\beta})' K_s (y - X\hat{\beta})}{r_s} \quad s = 0, 1, \dots, n-1,$$

so wird deutlich, daß die Lösungen $(\hat{\beta}, \hat{\lambda})$ dieser Gleichungen stets in jenen Unterräumen des Parameterraums liegen müssen, in denen die Beziehungen $(*)$ und $(**)$ jeweils für sich getrennt erfüllt sind:

$(*)$ liefert stationäre Punkte der Likelihoodfunktion für $\hat{\beta}$ bei gegebenem $\hat{\lambda}$.

$(**)$ liefert stationäre Punkte der Likelihoodfunktion für $\hat{\lambda}$ bei gegebenem $\hat{\beta}$.

Da beide Gleichungen zusammen nicht explizit nach den gesuchten Parametern auflösbar sind, muß ein Iterationsverfahren zur Lösung der Gleichungen angesetzt werden. Es liegt nun nahe, die Iterationsschritte so zu wählen, daß aufeinanderfolgende Iterierte jeweils im durch $(*)$ bzw. $(**)$ definierten Unter-
raum liegen. Der entsprechende Algorithmus lautet:

[1; Initialisierung]

Setze $i \leftarrow 0$.

Wähle eine konsistente Schätzung $\hat{\beta}^{(i)}$ für β , etwa

$$\hat{\beta}^{(i)} \leftarrow (X'X)^{-1} X'y$$

und berechne mittels $(**)$ die entsprechenden ASR-Schätzer für $s=0, 1, \dots, n-1$

$$\hat{\lambda}_s^{(i)} \leftarrow \frac{(y - X\hat{\beta}^{(i)})' K_s (y - X\hat{\beta}^{(i)})}{r_s}$$

[2; Neues $\hat{\beta}$]

Benutze $\hat{\lambda}_s^{(i)}$, um über (*) eine Schätzung vom AITKEN-Typ für β zu ermitteln:

$$\hat{\beta}^{(i+1)} = [X' \left[\sum_{s=0}^{n-1} \frac{1}{\hat{\lambda}_s^{(i)}} K_s \right] X]^{-1} X' \left[\sum_{s=0}^{n-1} \frac{1}{\hat{\lambda}_s^{(i)}} K_s \right] y$$

[3; Neues $\hat{\lambda}$]

Benutze $\hat{\beta}^{(i)}$, um über (**) eine Schätzung vom ASR-Typ für λ zu gewinnen:

$$\hat{\lambda}_s^{(i)} = \frac{(y - X\hat{\beta}^{(i)})' K_s (y - X\hat{\beta}^{(i)})}{r_s}$$

[4; Test]

Stop, sofern sich $(\hat{\beta}^{(i)}, \hat{\lambda}^{(i)})$ hinreichend wenig von $(\hat{\beta}^{(i+1)}, \hat{\lambda}^{(i+1)})$ unterscheidet.

Sonst setze $i \leftarrow i+1$ und $\rightarrow$ Schritt [2].

Betrachtet man diesen Algorithmus genauer, so stellt man fest, daß er unnütze Schritte enthält - produziert werden zwei Folgen $(\beta^{(0)}, \lambda^{(1)}, \beta^{(2)}, \lambda^{(3)}, \dots)$ und $(\lambda^{(0)}, \beta^{(1)}, \lambda^{(2)}, \beta^{(3)}, \dots)$, die in folgender Weise zusammenhängen:



Eine der beiden Folgen ist überflüssig, da aufgrund der Initialisierung $\lambda^{(0)}$ und $\lambda^{(1)}$ ASR-Schätzer sind, die mit dem gleichen $\beta^{(0)}$ gewonnen werden, die also identisch sind. Es genügt folglich, eine der beiden Folgen, etwa die Folge $(\beta^{(0)}, \lambda^{(1)}, \beta^{(2)}, \dots)$ zu berechnen.

Der Algorithmus entspricht genau dem Scoring-Verfahren von FISHER, wie man folgendermaßen erkennt:

Bezeichnen Σ_i, q_i, P_i, D_i die jeweils an der Stelle $(\beta^{(i)}, \lambda^{(i)})$ berechneten $\Sigma, (\frac{\partial S}{\partial \beta}), (\frac{\partial S}{\partial \lambda}), \text{diag}[\frac{r_s}{2\lambda_s}]$, so ist

$$\begin{aligned}\beta^{(i+1)} &= \beta^{(i)} + [X' \Sigma_i^{-1} X]^{-1} [X' \Sigma_i^{-1} y - X' \Sigma_i^{-1} X \beta^{(i)}] \\ &= \beta^{(i)} + [X' \Sigma_i^{-1} X]^{-1} q_i\end{aligned}$$

und

$$\lambda^{(i+1)} = \lambda^{(i)} + D_i^{-1} P_i ;$$

die Iterationsvorschrift lautet also

$$\begin{bmatrix} \beta^{(i+1)} \\ \lambda^{(i+1)} \end{bmatrix} = \begin{bmatrix} \beta^{(i)} \\ \lambda^{(i)} \end{bmatrix} + J_i^{-1} \begin{bmatrix} q_i \\ r_i \end{bmatrix}$$

wobei

$$J_i = \begin{bmatrix} X' \Sigma_i^{-1} X & 0 \\ 0 & D_i \end{bmatrix}$$

die an der Stelle $(\beta^{(i)}, \lambda^{(i)})$ genommene Informationsmatrix des Modells ist. Wegen positiver Definitheit von J_i ist das Verfahren offenbar akzeptabel (cf. BARD [1974]). Konvergenzuntersuchungen für derartige Gradientenverfahren führen nun zu Bedingungen (hier über die Eigenwerte von J_i^{-1}), die in der Praxis nur schwer zu verifizieren sind. Erfahrungen mit dem hier vorgestellten Modell zeigen jedoch, daß der Algorithmus rapide konvergiert (2 - 10 Iterationsschritte genügen in der Regel zur Stabilisierung aller Iterierten im Rahmen der Rechengenauigkeit; häufig liefern bereits $\lambda^{(1)}$ und $\beta^{(2)}$ ausgezeichnete Näherungen für die endgültigen Lösungen).

Da der Algorithmus dann zusammenbricht, wenn irgendein $\hat{\lambda}_s^{(i)}$ auf den Rand des Parameterraums gerät (also gleich 0 wird), sei eine hinreichende Bedingung angegeben, die eine solche Situation fast sicher verhindert. Die Bedingung ist nicht scharf (eine exaktere Bedingung findet sich im Beweis des Satzes), sichert jedoch gleichzeitig die Existenz von MINQU-Schätzern und ist in der Praxis leicht einzuhalten.

SATZ 4.2.3.1. : Falls die Zahl der Regressoren (k) kleiner ist als $\min(\frac{1}{2} r_0, \frac{1}{2} r_1, \dots, \frac{1}{2} r_{n-1})$, gilt für fast alle Beobachtungen y , daß alle Glieder der Folge $(\hat{\lambda}^{(0)}, \hat{\lambda}^{(1)}, \dots, \hat{\lambda}^{(i)}, \hat{\lambda}^{(i+1)}, \dots)$ im Inneren des Parameterraums liegen.

BEWEIS : $\lambda_s^{(0)}$ und $\lambda_s^{(1)}$ sind ungleich 0, sofern

$$y'(I - X(X'X)^{-1}X')K_s(I - X(X'X)^{-1}X')y \neq 0.$$

Für fast alle y ist dies der Fall, wenn

$$(I - X(X'X)^{-1}X')K_s(I - X(X'X)^{-1}X') \neq 0,$$

d.h. wenn

$$\text{tr } K_s \neq \text{tr } X'K_sX(X'X)^{-1}.$$

Nun ist

$$0 \leq \text{tr } X'K_sX(X'X)^{-1} \leq k,$$

so daß $\hat{\lambda}_s^{(0)}, \hat{\lambda}_s^{(1)}$ fast sicher positiv sind, sofern

$$k < r_i.$$

Sei nun $\lambda_s^{(i)} \neq 0$, dann ist $\lambda_s^{(i+1)} \neq 0$ für fast alle y , sofern

$$(I - \Sigma_i^{-1}X(X'\Sigma_i^{-1}X)^{-1}X')K_s(I - X(X'\Sigma_i^{-1}X)^{-1}X'\Sigma_i^{-1}) = 0.$$

Dies führt auf die Bedingung

$$\begin{aligned} \text{tr } K_s - 2 \text{tr } (X' \Sigma_i^{-1} X)^{-1} X' \left(\frac{1}{\lambda_s^{(i)}} K_s \right) X + \text{tr } (X' \Sigma_i^{-1} X)^{-1} \\ X' K_s X (X' \Sigma_i^{-1} X)^{-1} X' \Sigma_i^{-2} X > 0 \end{aligned}$$

Benutzen wir die Tatsache, daß die Spur der Matrix

$$C_s^{(i)} := (X' \Sigma_i^{-1} X)^{-1/2} X' \left(\frac{1}{\lambda_s^{(i)}} K_s \right) X (X' \Sigma_i^{-1} X)^{-1/2}$$

im Intervall $[0, k]$ liegt, so ist die äquivalente Bedingung

$$\text{tr } K_s + \sum_{t=0}^{n-1} \frac{\lambda_s^{(i)}}{\lambda_t^{(i)}} \text{tr } C_s^{(i)} C_t^{(i)} > 2 \text{tr } C_s^{(i)}$$

fast sicher erfüllt, wenn $k > \min \frac{1}{2} r_s$.

Unter der hinreichenden Bedingung des vorausgegangenen Satzes sind alle Fixpunkte $(\beta^{(i+1)} = \beta^{(i)} \text{ und } \lambda^{(i+1)} = \lambda^{(i)})$ des Algorithmus Lösungen der Likelihoodgleichungen, da J_i^{-1} unter dieser Bedingung stets existiert und ungleich 0 ist.

Betrachten wir noch den Fall vollständiger Suffizienz der Statistik $(y'k_0 y, \dots, y'K_{n-1} y, X'y)$. Es ist dort

$$\frac{\partial S}{\partial \beta} = (X' \Sigma^{-1} X) (X' X)^{-1} (X' y - X' X \beta)$$

wobei $(X' \Sigma^{-1} X) (X' X)^{-1}$ nichtsingulär ist. Die Likelihoodgleichungen sind damit explizit und eindeutig lösbar und führen auf die Schätzer

$$\begin{aligned} \hat{\beta} &= (X' X)^{-1} X' y \\ \hat{\lambda}_s &= \frac{(y - X \hat{\beta})' K_s (y - X \hat{\beta})}{\text{tr } K_s} \end{aligned}$$

Da in diesem Spezialfall der Erwartungswert von $\hat{\lambda}_s$ gleich

$$E[\hat{\lambda}] = \lambda_s \left(1 - \frac{\text{tr}(X' X)^{-1} X' K_s X}{\text{tr } K_s} \right)$$

ist, liefert eine einfache Biaskorrektur beste unverzerzte

Schätzer für λ_s :

$$\lambda_s^* = \frac{\hat{\lambda}_s}{\frac{\text{tr}(X'X)^{-1}X'K_sX}{1 - \frac{\text{tr} K_s}{\text{tr} K_s}}}$$

4.2.4. KONSISTENZ UND ASYMPTOTISCHE EFFIZIENZ
DER MAXIMUM-LIKELIHOOD-SCHÄTZER

Betrachten wir die Folge $(\hat{\beta}^{(0)}, \hat{\lambda}^{(1)}, \hat{\beta}^{(2)}, \hat{\lambda}^{(3)} \dots)$ der Iterierten, die vom eben vorgestellten Algorithmus erzeugt wird. Wir wissen, daß $\hat{\lambda}^{(1)}$ konsistent und asymptotisch effizient ist, und wir wissen weiter, daß mit einem $\hat{\lambda}^{(i)}$, das konsistent und asymptotisch effizient ist, auch $\hat{\beta}^{(i+1)}$ diese Eigenschaften besitzt. Es fehlt uns noch ein Satz, der aus den Eigenschaften von $\hat{\beta}^{(i+1)}$ auf die Eigenschaften von $\hat{\lambda}^{(i+2)}$ schließt, um für alle Glieder der Folge, beginnend mit $\hat{\lambda}^{(1)}$, Konsistenz und asymptotische Effizienz sicherzustellen.

SATZ 4.2.4.1. : Sei $\hat{\beta}$ eine Schätzfunktion für β , so daß $\sqrt{N}(\hat{\beta} - \beta)$ gegen einen $\mathcal{N}[0, V^{-1}]$ -verteilten Zufallsvektor mit $V^{-1} = \lim_{N \rightarrow \infty} (\frac{1}{N} X' \Sigma^{-1} X)^{-1}$ in Verteilung konvergiert. Dann konvergiert $\sqrt{r_s}(\hat{\lambda}_s - \lambda_s)$ in Verteilung gegen eine $\mathcal{N}[0, 2\lambda_s^2]$ -verteilte Zufallsvariable, wobei

$$\hat{\lambda}_s = \frac{(y - X\hat{\beta})' K_s (y - X\hat{\beta})}{\text{tr} K_s}.$$

Beweis : Es gilt die Zerlegung

$$\hat{\lambda}_s = \frac{u'K_s u}{r_s} + \frac{(\hat{\beta} - \beta)' X' K_s X (\hat{\beta} - \beta)}{r_s} + \frac{2 u' K_s X (\hat{\beta} - \beta)}{r_s}$$

Also ist

$$\begin{aligned} \sqrt{r_s} (\hat{\lambda}_s - \lambda_s) &= \left(\frac{u' K_s u}{\sqrt{r_s}} - \sqrt{r_s} \lambda_s \right) + \frac{\sqrt{N} (\hat{\beta} - \beta)' \left(\frac{1}{N} X' K_s X \right) \sqrt{N} (\hat{\beta} - \beta)}{\sqrt{r_s}} \\ &\quad + \frac{2 \frac{1}{\sqrt{N}} u' K_s X \sqrt{N} (\hat{\beta} - \beta)}{\sqrt{r_s}} \end{aligned}$$

Betrachten wir die drei Terme getrennt:

$$(1) \quad E \left[\frac{u' K_s u}{\sqrt{r_s}} \right] = \sqrt{r_s} \lambda_s$$

$$\text{var} \left[\frac{u' K_s u}{\sqrt{r_s}} \right] = 2 \frac{\text{tr} K_s \Sigma K_s \Sigma}{r_s} = 2 \lambda_s^2$$

Für höhere Kumulanten $\kappa_h \left(\frac{u' K_s u}{\sqrt{r_s}} \right)$ ($h > 2$) findet man, daß

$$\kappa_h \left(\frac{u' K_s u}{\sqrt{r_s}} \right) \rightarrow 0 \quad \text{für } r_s \rightarrow \infty$$

und folglich

$$\frac{u' K_s u}{\sqrt{r_s}} - \sqrt{r_s} \lambda_s \xrightarrow{V} \mathcal{W}[0, 2\lambda_s^2]$$

(2) Der zweite Term konvergiert stochastisch gegen 0 ,

da $\left(\frac{1}{N} X' K_s X \right)$ gegen die feste Matrix V_s konvergiert; der Zähler des 2. Terms folglich nach Verteilung gegen eine quadratische Form eines normalverteilten Zufallsvektors konvergiert.

(3) Es ist schließlich für jedes N

$$X'K_s u \sim \mathcal{W}[0, X'K_s X]$$

folglich

$$\frac{1}{N} X'K_s u \sim \mathcal{N}[0, \frac{1}{N} X'K_s X] \xrightarrow{V} \mathcal{N}[0, V_s]$$

wobei V_s eine feste endliche Matrix ist.

Auch der dritte Term konvergiert also stochastisch gegen 0, womit alles gezeigt ist.

Ein alternativer Ansatz zur Untersuchung der asymptotischen Eigenschaften betrachtet nicht die vom Algorithmus zur Lösung der Likelihoodgleichungen erzeugten Iterierten, sondern direkt die absoluten Maxima der Likelihoodfunktion (siehe etwa HARTLEY, RAO [1967]).

Anschaulich gesprochen, beruht die Vorgehensweise auf zwei Eigenschaften der Likelihoodfunktion, die in "regulären" Situationen erfüllt sind:

- (1) Der Erwartungswert $E_0[\frac{1}{N} S(\theta_0 | y)]$ der Loglikelihood an der "wahren" Parameterstelle ist echt größer als der Erwartungswert $E_0[\frac{1}{N} S(\theta | y)]$ an jeder anderen Stelle.
- (2) Die beobachtete Loglikelihood $\frac{1}{N} S(\theta | y)$ konvergiert stark [WALD, 1949] bzw. stochastisch [WOLFOWITZ, 1949] gegen den jeweiligen Erwartungswert $E_0[\frac{1}{N} S(\theta | y)]$.

Es ist dann zu erwarten, daß "in großen Stichproben" die "absoluten" ML-Schätzer sich in der Nähe des wahren Parameterpunkts konzentrieren werden.

In unserem Fall wurde (1) bereits gezeigt; Konvergenz im Mittel von $\frac{1}{N} S(\theta | y)$ folgt aus der Tatsache, daß wegen

$$y - XB \sim \mathcal{N}[X(\beta_0 - B), \Sigma_0]$$

der Ausdruck

$$\text{var } \frac{1}{N} S(\theta | y) = \frac{1}{N^2} \frac{1}{4} \text{var}[(y - X\beta)' \Sigma^{-1} (y - X\beta)]$$

$$= \frac{1}{N^2} \frac{1}{4} [2 \text{tr} \Sigma^{-1} \Sigma_o \Sigma^{-1} \Sigma_o + 4(\beta - \beta_o)' X' \Sigma^{-1} \Sigma_o \Sigma^{-1} X (\beta_o - \beta)]$$

ein $O(\frac{1}{N})$ ist.

5. VERGLEICH DER SCHÄTZFUNKTIONEN

5.1. ZUSAMMENSTELLUNG DER ERGEBNISSE

Die untersuchten statistischen Eigenschaften der Schätzfunktionen für die Eigenwerte der Kovarianzmatrix erlauben keine eindeutige Diskrimination. Je nach Verwendungszweck wird

- Unverzerrtheit
- Varianzminimalität an einem bestimmten Parameterpunkt
- Einfachheit der Berechnung
- Möglichkeit, die exakte Verteilung leicht angeben zu können
- Maximierung der Likelihood

das für die Auswahl eines bestimmten Schätzers entscheidende Kriterium sein.

Die folgende Tabelle faßt die Ergebnisse der letzten Kapitel zusammen - dabei werden neben den dort untersuchten Eigenschaften noch zwei etwas schwieriger zu operationalisierende Kriterien mit aufgenommen: der Rechenaufwand (der wesentlich von den verfügbaren Programmen abhängt) und die anschauliche Begründbarkeit der Schätzfunktionen (ein wichtiges Kriterium für jeden, der statistisch unvorbelastete Anwender zu beraten hat). Alle betrachteten Schätzfunktionen sind beste asymptotisch normalverteilte Schätzer und damit asymptotisch identisch - es werden daher in der Tabelle nur finite Eigenschaften aufgeführt.

	ASR	GEN.ASR	MINQU	GEN.MINQU	BALESTRA	ML
Unverzerrtheit	nein	nein	ja	ja	ja	nein
Translationsinvarianz	ja	ja	ja	ja	ja	ja
Nichtnegativität	ja	ja	nicht notwendig	nicht notwendig	ja	ja
Existenz sofern $\text{tr } K_s > 0$	ja	ja	nicht notwendig	nicht notwendig	nicht notwendig	nicht notwendig
Maximierung der Likelihood	nein	nein	nein	nein	nein	ja
Lokale Varianzminimalität	nein	nein	ja	ja	nein	nein
Exakte Chiquadratverteilung	nein	nein	nein	nein	ja	nein
Rechenaufwand	gering	gering	mittel	mittel	gering bis mittel	hoch
Anschauliche Begründbarkeit	ja	nein	nein	nein	ja	ja

Da lokale Varianzminimalität ein sehr schwaches Kriterium darstellt und da in den seltensten Fällen Vorkenntnisse über die Parameter λ vorliegen (um generalisierte ASR-Schätzer bilden zu können), sind für die Praxis vor allem 3 Schätzfunktionen von Interesse: ASR-Schätzer, ML-Schätzer, BALESTRA-Schätzer.

ASR-Schätzer sind anschaulich gut zu begründen, sehr leicht zu berechnen, existieren stets sofern $\text{tr } K_s > 0$, sind translationsinvariant und positiv. Die Verzerrtheit dieser Schätzer ist in "großen" Stichproben unerheblich - in kleinen Stichproben kann leicht die Abschätzung für die Größe des Bias durch

Betrachtung der Matrizen $X'K_sX$ gewonnen werden.

ML-Schätzer sind vor allem aufgrund ihrer Invarianzeigenschaft wichtig, sofern einmal das Likelihoodprinzip akzeptiert wird. Sie teilen ihre "Maximum-Likelihood"-Eigenschaft allen aus ihnen abgeleiteten Statistiken mit, d.h. alle Funktionen der ML-Schätzer für λ sind gleichfalls ML-Schätzer für die entsprechenden Funktionen von λ (ZEHNA [1966]). Diese Eigenschaft ist wesentlich, falls die Schätzfunktionen nicht nur zur Beurteilung der Variabilität von β benutzt werden, sondern Grundlage für inhaltlich zu interpretierende Kenngrößen der Beobachtungen bilden sollen.

BALESTRA-Schätzer schließlich sind diejenigen Statistiken, die sich dem an finiten Tests interessierten Anwender aufgrund ihrer exakten Chiquadratverteilung aufdrängen. Sie sind zwar nicht die "besten" quadratisch unverzerrten Schätzer, aber sicher recht "gute" Schätzer (Unverzerrtheit, Translationsinvarianz, Positivität, Anschaulichkeit). Sofern alle Matrizen $X'K_sX$ existieren, sind sie zudem mit Hilfe von Standardsoftwarepaketen wie SPSS leicht zu berechnen.

5.2. SKIZZE EINIGER SIMULATIONSERGEBNISSE

Ein tiefgehenderer Vergleich der Schätzfunktionen setzt Simulationsstudien voraus. Dies aus zwei Gründen:

- (1) Eine Reihe von Ergebnissen sind auf analytischem Wege überhaupt nicht zu erzielen (so liegen beispielsweise keine Informationen über finite Eigenschaften der ML-Schätzer vor)
- (2) Auch dort, wo analytische Resultate existieren, sind diese häufig wenig instruktiv (siehe etwa die Varianz-

Kovarianzmatrix von ASR- oder MINQU-Schätzern).

Simulationsstudien im Bereich der Theorie linearer Modelle sind nun in der Regel weit zeitaufwendiger als vergleichbare Untersuchungen in anderen Gebieten der Statistik: Die große Zahl von "nuisance"-Parametern (Regressormatrizen X , Mehrebenenstruktur in $K_0, K_1, \dots, K_{n-1}$) erzwingt Simulationsläufe für sehr viele verschiedene $(X, K_0, \dots, K_{n-1})$ -Konfigurationen, um auch nur einigermaßen allgemeingültige Resultate zu sichern. Derart umfassende Simulationsexperimente konnten aufgrund der hohen Störanfälligkeit der verwendeten Rechananlage nicht durchgeführt werden. Die im folgenden kurz skizzierten Ergebnisse für den Zweiebenenfall sind daher wegen der geringen Zahl (10 - 20) der jeweils zugrundegelegten Konfigurationen nur als Illustrationen zu werten. (Anm. 1).

(1) Das Verhalten der Schätzer bei Abweichungen von der Normalverteilung

Maximum-Likelihood-Schätzer $(\hat{\beta}, \hat{\lambda})$ und Ausdrücke für die Varianzen der Schätzfunktionen für λ wurden unter Normalverteilungsannahmen gewonnen, so daß es nahe liegt, die Robustheit dieser Ergebnisse zu untersuchen, wenn alternative Verteilungen für die Erzeugung der Störungen benutzt werden.

Anm. 1: Um Unterschiede zwischen den (asymptotisch äquivalenten) Schätzern zu erkennen, wurden Stichproben benutzt, die kleiner sind als die üblicherweise in der Sozialforschung anzutreffenden ($N=100$, $m=20$, $k=2$). Für jede Konstellation (bestehend aus Matrix der Regressoren, Zerlegung $\mathcal{Z}$ und Parameterwerten) wurden jeweils 100 Replikationen gerechnet - bei weiteren Läufen traten nur geringfügige Veränderungen der Ergebnisse ein.

Die üblicherweise durchgeführten Robustheitsstudien (siehe etwa die faszinierende Princeton Robustness Study - ANDREWS, D.F./.../TUKEY, J.W. [1972]) sind durchgängig für sozialwissenschaftliche Anwendungen nur von geringer Relevanz. Dies liegt daran, daß in derartigen Untersuchungen das Verhalten statistischer Methoden bei Daten aus Verteilungen mit sehr langen "tails" studiert wird. Sozialwissenschaftliche Variable, die in der Regel durch Kombination von ordinalskalierten Fragen mit wenigen Kategorien entstehen, haben nun von vornherein beschränkte Wertebereiche und damit kurze "tails": Probleme mit "Ausreißern" können dort nicht auftreten.

Es wurden hier als alternative Verteilungen für die Störgrößen die Klasse der Betaverteilungen gewählt (Beschränktheit, große Vielfalt von Verteilungsformen) - also Zufallsvariablen mit der Dichtefunktion

$$f(x) = \begin{cases} \frac{1}{B(p,q)} \frac{(x-a)^{p-1} (b-x)^{q-1}}{(b-a)^{p+q-1}} & a \leq x \leq b \\ 0 & \text{sonst} \end{cases} \quad (\text{Anm. 1})$$

Der Bereich $[a,b]$ wurde dabei durch die Forderung festgelegt, daß bei vorgegebenen Parametern p,q

$$E[X] = 0 \quad \text{und} \quad \text{var}[X] = 1$$

gelten sollte. Ein Vektor e von derartigen unabhängig beta-verteilten Zufallszahlen (zum Erzeugungsalgorithmus siehe JÖHNK [1964]) kann dann durch die lineare Transformation

$$u = (\sqrt{\lambda_0} K_0 + \sqrt{\lambda_1} K_1) e$$

in einen Zufallsvektor mit der gewünschten Varianz-Kovarianz-

Anm. 1: Dabei ist $p,q > 0$ und $B(p,q) = \int_0^1 t^{p-1} (1-t)^{q-1} dt$
 die Betafunktion.

matrix überführt werden.

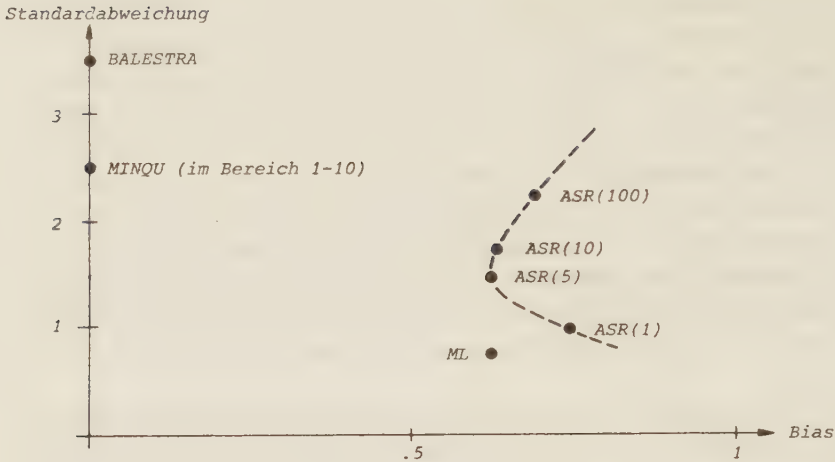
Die Resultate der (noch sehr vorläufigen) Simulationsläufe sind überraschend:

- (a) Die relative Lage der Schätzer für λ_1 zueinander ist im wesentlichen die gleiche wie unter Normalverteilungsannahmen, was Bias und Varianz betrifft (siehe später unter Punkt 2). Maximum-Likelihood-Schätzer liegen dabei sogar relativ etwas günstiger als im Normalverteilungsfall - ihre Varianz ist geringfügig niedriger.
- (b) Bei Schätzern für β und λ_0 ist kein Unterschied festzustellen (dies ist wohl auf die höhere Zahl an Freiheitsgraden zurückzuführen, auf denen Schätzfunktionen für diese Parameter basieren).
- (c) Die Varianzen aller Schätzer für λ_1 waren durchweg etwas niedriger als in den hinsichtlich X-Konfigurationen, Gruppengrößen und Populationsparametern vergleichbaren Läufen mit normalverteilten Störungen.

Diese Ergebnisse, die auch für sehr extreme Alternativen wie U-förmige und schiefe Verteilungen zutreffen, sind m.E. dadurch zu erklären, daß standardisierte Beta-Verteilungen kürzere "tails" als die Standardnormalverteilung aufweisen - die empirischen Varianzen und Kovarianzen der Störungen sind folglich enger um die entsprechenden Populationsparameter konzentriert.

(2) Schätzer für λ_1

Die typische Konfiguration der Schätzfunktionen für λ_1 in einem (Bias, Standardabweichung)-Diagramm sei an folgender Abbildung illustriert (Populationsparameter $\lambda_0 = 1$, $\lambda_1 = 5$)



Die (generalisierten) ASR-Schätzer wurden durch die jeweiligen A-priori-Annahmen über das Verhältnis λ_1/λ_0 der Eigenwerte parametrisiert. Das Vorzeichen des deutlich zu erkennenden Bias dieser Schätzer ist abhängig von der jeweils zugrunde gelegten X-Konstellation (siehe Satz 3.2.2.2.). Optimal in der Klasse der GASR-Schätzer sind, wie zu vermuten war, Schätzer, die unter zutreffenden A-priori-Annahmen gewonnen wurden (hier ASR(5)) - der Effekt falscher A-priori-Annahmen ist allerdings innerhalb einer großen Bandbreite (1 - 100) derartiger Annahmen relativ gering.

Maximum-Likelihood-Schätzer liegen durchweg günstiger als ASR-Schätzer: ihr Bias ist gleich dem Bias der optimalen ASR-Schätzer, ihre Varianz jedoch geringer.

MINQU-Schätzer sind unverzerrt und es ist im wesentlichen irrelevant, welche A-priori-Annahmen über λ_1/λ_0 zu ihrer Ableitung benutzt werden (dies gilt allerdings nicht für MINQU-Schätzer

mit $\lambda_1/\lambda_0 \ll 1$.) Sie sind, was Standardabweichungen betrifft, den PALESTRA-Schätzern überlegen - eine Illustration der Tatsache, daß diese Schätzer weder lokale noch globale Optimalitätseigenschaften aufweisen.

(3) Schätzer für β und λ_0

Auch unter Normalverteilungsannahmen scheint es relativ gleichgültig, welche Schätzer für β und λ_0 gewählt werden, da wegen der vergleichsweise großen Zahl an Freiheitsgraden, auf denen diese Schätzer beruhen, deren asymptotische Äquivalenz früher erkennbar wird als bei Schätzern für λ_1 .

Die Simulationsläufe bestätigen jedoch die Untersuchungen in Kapitel 2, nach denen die konventionelle "OLS-Schätzung" für die Varianz-Kovarianzmatrix von β^* selbst bei schwach korrelierten Beobachtungen zu bedenklichen Unterschätzungen der "Standardfehler" der Regressionskoeffizienten führt.

6. MÖGLICHE ERWEITERUNGEN

6.0. VORBEMERKUNG

Im Hauptteil der Arbeit wurden Schätzverfahren für das univariate "Mehrebenenmodell" untersucht. Für die praktische Anwendung in der Sozialforschung müssen darüber hinaus allerdings weitere Probleme gelöst werden:

- (1) Die Ableitung von Schätzfunktionen für das multivariate Modell - multivariate Regressionsverfahren werden gerade in Mehrebenenstudien häufig verwendet (siehe HARDER/PAPPI [1969]).
- (2) Die Entwicklung von Tests für Hypothesen über die Regressionskoeffizienten β und Hypothesen über λ (insbesondere Tests auf Unabhängigkeit der Störungen).
- (3) Die Überprüfung des Modells selbst.

Mögliche Richtungen für die Weiterarbeit an diesen Problemen werden im folgenden kurz angedeutet.

6.1. MULTIVARIATE REGRESSIONSMODELLE

Das Modell ist

$$Y = XB + U$$

mit

$$\text{var}[\text{vec } U] = \Sigma = K_0 \otimes \Omega_0 + K_1 \otimes \Omega_1 + \dots + K_{n-1} \otimes \Omega_{n-1} \quad .$$

Die Ableitung von Schätzern für die "Kreuzebenenkovarianzmatrizen" Ω_s und die Untersuchung deren finiter und asymptotischer Eigenschaften können hier ganz wie im univariaten Modell geführt werden - an die Stelle des grundlegenden Lemmas 3.2.2.1. tritt dabei ein entsprechender Satz, der auf die Möglichkeit beruht, Matrizen $Y'AY$ bilinearer und quadratischer Formen normalverteilter Spalten von Y als Summe von nichtzentral WISHART-verteilten Zufallsmatrizen darzustellen.

Die Schätzfunktionen für (B, Ω_s) sind analog zu den bisher betrachteten Schätzern für (β, λ_s) - als Illustration sei der Algorithmus zur Lösung der Likelihood-Gleichungen des multivariaten Modells genannt, der in der abwechselnden Iteration von

$$(*) \text{vec } \hat{B} = \left[\sum_{s=1}^m X' K_s X \hat{\Omega}_s^{-1} \right]^{-1} \left[\sum_{s=1}^m X' K_s \hat{\Omega}_s^{-1} \right] \text{vec } Y$$

und $(**) \hat{\Omega}_s = (\text{tr } K_s)^{-1} (Y - XB)' K_s (Y - XB)$

besteht.

6.2. TESTPROBLEME FÜR λ UND β

In sehr approximativer Betrachtung kann die asymptotische Normalverteilung der Schätzer für λ und β zur Grundlage für asymptotische Tests gewählt werden. Angesichts der in der Mehrebenenanalyse üblichen Stichprobengrößen dürften diese Tests häufig bereits recht zufriedenstellend sein.

Mit den exakten chiquadratverteilten BALESTRA-Schätzern $\lambda_s^{(B)}$ ist es möglich, Tests auch in kleineren Stichproben zu gewinnen. Die Vorgehensweise sei für den Zweiebenenfall skizziert:

6.2.1. TEST AUF UNABHÄNGIGKEIT DER STÖRUNGEN

Bei unabhängigen Störungen gilt $\lambda_1 - \lambda_0 = 0$.

Das Testproblem

$$H_0 : \lambda_1 = \lambda_0$$

versus

$$H_1 : \lambda_1 \neq \lambda_0$$

führt wegen

$$d_s \frac{\lambda_s^{(B)}}{\lambda_s} \sim \mathcal{Q}[d_s]$$

mit

$$d_s = \text{tr } K_s - \text{rg } X'K_sX \quad (s = 0, 1)$$

auf die naheliegende Prüfgröße

$$\Phi = \frac{\frac{d_0 \frac{\lambda_0^{(B)}}{\lambda_0} / d_0}{\frac{d_1 \frac{\lambda_1^{(B)}}{\lambda_1} / d_1}} = \frac{\lambda_0^{(B)}}{\lambda_1^{(B)}} \quad (\text{bei } \lambda_0 = \lambda_1)$$

die unter H_0 zentral F-verteilt ist mit d_0 und d_1 Freiheitsgraden.

6.2.2. TESTS FÜR DIE REGRESSIONSKOEFFIZIENTEN

Das allgemeine lineare Testproblem für β lautet

$$H_0 : R'\beta = m$$

versus

$$H_1 : R'\beta \neq \pi$$

Dabei ist R' eine vorgegebene (s,k) -Matrix mit vollem Zeilenrang s und m ein vorgegebener $(s,1)$ -Vektor. Bei bekanntem Σ ist $\hat{\beta} = (X'\Sigma^{-1}X)^{-1}X'\Sigma^{-1}y$ eine beste unverzerrte Schätzfunktion für β und wegen

$$R'\hat{\beta} - m \sim N[R'\beta - m, R'(X'\Sigma^{-1}X)^{-1}R]$$

kann ein Test der allgemeinen linearen Hypothese auf der Tatsache aufgebaut werden, daß unter H_0 die quadratische Form

$$Q = (R'\hat{\beta} - m)'(R'(X'\Sigma^{-1}X)^{-1}R)^{-1}(R'\hat{\beta} - m)$$

zentral chiquadratverteilt ist mit s Freiheitsgraden. Es gilt dann also

$$P[Q \leq \chi^2_{s;1-\alpha}] = 1 - \alpha$$

wobei α das vorgegebene "Signifikanzniveau" des Tests und $\chi^2_{s;1-\alpha}$ der $(1-\alpha)\%$ -Prozentpunkt der $C[s]$ -Verteilung ist.

Der Ansatz von WELSH [1947] und JAMES [1951, 1954] zur approximativen Lösung des BEHRENS-FISHER-Problems führt hier nun darauf, bei unbekannten λ_0 und λ_1 die unverzerrten Schätzer $\lambda_0^{(B)}$ und $\lambda_1^{(B)}$ zu benutzen und nach einer Funktion $h(\lambda_0^{(B)}, \lambda_1^{(B)})$ zu suchen, so daß in möglichst guter Näherung die Beziehung

$$P[Q(\lambda_0^{(B)}, \lambda_1^{(B)}) \leq h(\lambda_0^{(B)}, \lambda_1^{(B)})] = 1 - \alpha$$

erfüllt ist (hier ist $Q(\lambda_0^{(B)}, \lambda_1^{(B)})$ die gleiche Funktion von $\lambda_0^{(B)}$ und $\lambda_1^{(B)}$ wie Q von λ_0 und λ_1).

Eine Taylorentwicklung von $h(\cdot, \cdot)$ bis zur ersten Ordnung d_0^{-1}, d_1^{-1} in den Freiheitsgraden d_0, d_1 der Schätzer für die Varianzen eines heteroskedastischen Modells wurde von JAMES [1954; S. 25] gegeben - mit den Beziehungen von Abschnitt 1.3. läßt sich diese Entwicklung auch auf das hier betrachtete Modell anwenden.

6.3. ÜBERPRÜFUNG DES MODELLS

Das vorgestellte Modell macht zwar soziologisch plausible, aber recht stringente Annahmen über die spezifische Art und Weise, in der die Kovarianzen zwischen Beobachtungen von der Struktur der Mehrebenen-situation abhängen. Datenanalytische Verfahren, um Einsicht in die Gültigkeit dieser Annahmen bei einer konkreten Datenlage zu gewinnen, wurden in STREITBERG [1976] vorgeschlagen. Ein formaleres Vorgehen beruht auf der Möglichkeit, das Modell in eine umfassende Modellklasse einzubetten, in der möglicherweise verschiedene Intragruppenkovarianzen in den einzelnen Kollektiven zugelassen werden.

Wir betrachten eine Zweiebenenstruktur $(\mathcal{U}, \mathcal{Z})$ und nehmen folgende Gestalt der Kovarianzmatrix in der r -ten Gruppe $\mathcal{U}_r$ an:

$$\Sigma_r = \begin{bmatrix} \rho_0 + \rho_r & \rho_r & \dots & \rho_r \\ \rho_r & \rho_0 + \rho_r & \dots & \rho_r \\ \cdot & \cdot & & \cdot \\ \cdot & \cdot & & \cdot \\ \rho_r & \rho_r & \dots & \rho_0 + \rho_r \end{bmatrix}$$

Dabei ist ρ_r ein gruppenspezifischer Parameter; ρ_0 ein Parameter, der die Varianz der individuellen Störungen bestimmt und über die Gruppen hinweg konstant ist.

Die gesamte Kovarianzmatrix ist dann

$$\Sigma = \Sigma_1 \oplus \Sigma_2 \oplus \dots \oplus \Sigma_r \oplus \dots \oplus \Sigma_m$$

Mit Hilfe eines kleinen Kunstgriffs kann diese Kovarianzmatrix

als Linearkombination idempotenter, wechselseitig orthogonaler Matrizen geschrieben werden.

Wir definieren eine (N, N) -Matrix M_r durch

$$M_r = (m_{ij}^{(r)}) = \begin{cases} 0 & \text{für } i \notin \mathcal{M}_r \text{ oder } j \notin \mathcal{M}_r \\ \frac{1}{n_r} & \text{für } i \in \mathcal{M}_r \text{ und } j \in \mathcal{M}_r \end{cases}.$$

Es gilt

$$M_r M_s = \begin{cases} 0 & \text{für } r \neq s \\ M_r & \text{für } r = s \end{cases}$$

Die Kovarianzmatrix Σ ist dann

$$\Sigma = \rho_0 I + n_1 \rho_1 M_1 + \dots + n_r \rho_r M_r + \dots + n_m \rho_m M_m.$$

Da die Matrix

$$(I - \Delta) = I - (M_1 + M_2 + \dots + M_m)$$

idempotent und orthogonal zu allen M_s ist, lautet die Spektraldarstellung von Σ :

$$\begin{aligned} \Sigma &= \rho_0 (I - \Delta) + (n_1 \rho_1 + \rho_0) M_1 + \dots + (n_m \rho_m + \rho_0) M_m \\ &= \rho_0 (I - \Delta) + \lambda_{11} M_1 + \dots + \lambda_{1m} M_m. \end{aligned}$$

Das Modell ist formal identisch mit dem in Kapitel 2 beschriebenen. Schätzfunktionen lassen sich daher mit den in Kapitel 3 und 4 dargestellten Techniken gewinnen. Die Schätzer für λ_{1r} beruhen wegen $\text{tr}(M_r) = 1$ jedoch nur auf einem Freiheitsgrad und sind mithin sehr instabil.

Die in Kapitel 2 - 5 untersuchten Zweiebenenmodelle mit den Parametern $(\beta, \lambda_0, \lambda_1)$ bilden eine Teilmenge des umfassenden

Modells, die durch die $(m-1)$ Restriktionen

$$r = (r_1, \dots, r_{m-1})$$

gekennzeichnet ist. Im Rahmen eines Likelihood-Ansatzes gibt es verschiedene Möglichkeiten, Tests auf Vorliegen des restringierten Modells (*) zu konstruieren:

- (1) Generalisierte Likelihoodratio-tests (man wählt das Verhältnis der Suprema der Likelihood-Funktion im restringierten Modell (*) und im unrestringierten Modell (**) als Prüfgröße).
- (2) Das Verfahren von WALD [1943], in dem geprüft wird, "wie gut" die ML-Schätzer im unrestringierten Modell die Restriktion (*) erfüllen.
- (3) Der "Lagrange-Multiplier"-Test von AITCHISON/SILVEY [1958, 1960] und SILVEY [1959], der es erlaubt, im restringierten Modell zu untersuchen, "wie weit" die restringierten ML-Schätzer vom absoluten Maximum der Likelihood-Funktion von (**) entfernt sind.

Alle drei Verfahren setzen jedoch eine gesonderte Untersuchung der Verteilung der jeweils benutzten Prüfgröße voraus, da die Annahmen, auf denen die asymptotische Theorie dieser Tests beruhen, im unrestringierten Modell verletzt sind.

7. LITERATURVERZEICHNIS

AITCHISON, J./SILVEY, S.D. [1958] MAXIMUM LIKELIHOOD
ESTIMATION OF PARAMETERS SUBJECT TO RESTRAINTS,
AMS 29 P:813-828

AITCHISON, J./SILVEY, S.D. [1960] MAXIMUM LIKELIHOOD
ESTIMATION PROCEDURES AND ASSOCIATED TESTS OF
SIGNIFICANCE, J.ROY.STAT.SOC.B 22 P:154-171

ALKER, HAYWARD R. [1969] A TYPOLOGY OF ECOLOGICAL
FALLACIES, IN: DOGAN/ROKKAN P:69-86

ANDREWS, D.F./BICKEL, P.J./HAMPEL, F.R./HUBER, P.J./ROGERS,
W.H./TUKEY, J.W. [1972] ROBUST ESTIMATES OF
LOCATION, PRINCETON UNIVERSITY PRESS PRINCETON

BALESTRA, P. [1973] BEST QUADRATIC UNBIASED
ESTIMATORS OF THE VARIANCE-COVARIANCE MATRIX
IN NORMAL REGRESSION, JOURNAL OF ECONOMETRICS 1
P:17-28

BARD, YONATHAN [1974] NONLINEAR PARAMETER ESTIMATION,
ACADEMIC PRESS NEW YORK

BARTON, A.H. [1968] BRINGING SOCIETY BACK IN: SURVEY-
RESEARCH AND MACRO-METHODOLOGY, AMERICAN
BEHAVIORAL SCIENTIST P:1-9

BELLMAN, RICHARD [1960] INTRODUCTION TO MATRIX
ANALYSIS, MC GRAW HILL NEW YORK

BOYD, L.H. [1971] MULTIPLE LEVEL ANALYSIS WITH
COMPLETE AND INCOMPLETE DATA, UNIV. OF MICHIGAN
DISSERTATION

DOGAN, MATTEI/ROKKAN, STEIN [1969] QUANTITATIVE
ECOLOGICAL ANALYSIS IN THE SOCIAL SCIENCES, THE
MIT PRESS CAMBRIDGE

GRAYBILL, F.A./HULTQUIST, R.A. [1961] THEOREMS
CONCERNING EISENHART'S MODEL 2, AMS 32 P:261-269

HALPERIN, MAX [1951] NORMAL REGRESSION THEORY IN THE
PRESENCE OF INTRA-CLASS CORRELATION, AMS 22
P:573-580

HARDER, THEODOR/PAPPI, FRANZ URBAN [1969] MULTIPLE-
LEVEL REGRESSION ANALYSIS OF SURVEY AND
ECOLOGICAL DATA, SOCIAL SCIENCE INFORMATION 8
P:43-67

HARTLEY, H.O./RAO, J.N.K. [1967] MAXIMUM LIKELIHOOD
ESTIMATION FOR THE MIXED ANALYSIS OF VARIANCE
MODEL, BIOMETRIKA 54 P:93-108

HUMMELL, H.J. [1972] PROBLEME DER MEHREBENENANALYSE,
TEUBNER STUTTGART

JAMES, G.S. [1951] THE COMPARISON OF SEVERAL GROUPS
OF OBSERVATIONS WHEN THE RATIOS OF THE
POPULATION VARIANCES ARE UNKNOWN, BIOMETRIKA 38
P:324-329

JAMES, G.S. [1954] TESTS OF LINEAR HYPOTHESES I.
UNIVARIATE AND MULTIVARIATE ANALYSIS WHEN THE
RATIOS OF POPULATION VAR. ARE UNKNOWN,
BIOMETRIKA 41 P:19-43

- JOHNSON, NORMAN L./KOTZ, SAMUEL [1970] CONTINUOUS UNIVARIATE DISTRIBUTIONS - 2, HOUGHTON MIFFLIN, BOSTON
- LAMOTTE, LYNN ROY [1973] ON QUADRATIC ESTIMATION OF VARIANCE COMPONENTS, BIOMETRICS 29
- LANCASTER, H.O. [1965] THE HELMERT MATRICES, AMERICAN MATHEMATICAL MONTHLY 72 P:4-11
- LEHMANN, E.L. [1959] TESTING STATISTICAL HYPOTHESES, WILEY NEW YORK
- MACDUFFEE, C.C. [1946] THE THEORY OF MATRICES, CHELSEA PUBL.CO. NEW YORK
- MACRAE, E. [1974] MATRIX DERIVATIVES WITH AN APPLICATION TO AN ADAPTIVE LINEAR DECISION PROBLEM, ANNALS OF STATISTICS 2 P:337-346
- RAO, C.R. [1972] ESTIMATION OF VARIANCE AND COVARIANCE COMPONENTS IN LINEAR MODELS, JASA 67 P:112-115
- RAO, C.R. [1973] LINEAR STATISTICAL INFERENCE AND ITS APPLICATIONS, WILEY NEW YORK
- SEARLE, S.R. [1971] LINEAR MODELS, WILEY NEW YORK

SILVERMAN, J.D. [1959] THE LAGRANGIAN MULTIPLIER TEST,
AMS 30 P:389-407

STREITBERG, BERND H.J. [1976] DEPENDENT OBSERVATIONS
IN MULTIPLE LEVEL ANALYSIS, COMPSTAT
PROCEEDINGS IN COMPUTATIONAL STATISTICS
PHYSICA VERLAG WIEN

WALD, ABRAHAM [1943] TESTS OF STATISTICAL HYPOTHESES
CONCERNING SEVERAL PARAMETERS WHEN THE NUMBER
OF OBSERVATIONS IS LARGE, TRANS. AMERICAN MATH.
SOCIETY 54 P:426-482

WALD, ABRAHAM [1949] NOTE ON THE CONSISTENCY OF THE
MAXIMUM LIKELIHOOD ESTIMATE, AMS 20 P:595-601

WELCH, B.L. [1947] THE GENERALIZATION OF STUDENT'S
PROBLEM WHEN SEVERAL DIFFERENT POPULATION
VARIANCES ARE INVOLVED, BIOMETRIKA 34 P:28-35

WOLFOWITZ, J. [1949] ON WALD'S PROOF OF THE
CONSISTENCY OF THE MAXIMUM LIKELIHOOD ESTIMATE,
AMS 20 P:601-602

ZEHNA, PETER W. [1966] INVARIANCE OF MAXIMUM
LIKELIHOOD ESTIMATORS, AMS 37 P:744

L E B E N S L A U F

- 1946 Am 18.12.46 als Sohn von Robert und Erna Streitberg in Bad Kissingen (Bayern) geboren.
- 1953 - 1957 Volksschule in Bad Kissingen.
- 1957 - 1966 Neusprachl. Gymnasium in Bad Kissingen.
- 1966 - 1967 Studium der Volkswirtschaftslehre und Soziologie an der Julius-Maximilian Universität, Würzburg.
- 1967 - 1972 Studium der Volkswirtschaftslehre, Soziologie und Mathematik an der FU Berlin. Abschluß als Diplom-Soziologe SS 1972. Seit SS 1972 Aufbau-studium der Mathematik an der FU Berlin.
- 1972 - 1974 Teilzeitassistent für empirische Sozialforschung und Statistik am Institut für Soziologie (FB 11, FUB).
- 1974 - 1976 Wiss. Assistent für Computational Statistics am Institut für Angewandte Statistik und Quantitative Ökonomie (FB 10, FUB).

